DATA ETHICS OF POWER
A Human Approach to Big Data and AI

Gry Hasselbalch
DATA ETHICS
OF POWER

A Human Approach to Big Data and AI

Gry Hasselbalch

University of Copenhagen
Faculty of Humanities
PhD thesis submitted at University of Copenhagen, Faculty of Humanities, January 2021.

Data Ethics of Power: A Human Approach to Big Data and AI
Copyright © 2021 by Gry Hasselbalch

Citation: Hasselbalch, G. (2021) *Data Ethics of Power: A Human Approach to Big Data and AI*, University of Copenhagen, Faculty of Humanities.
To Clara & Francesco
ABSTRACT

Data Ethics of Power – A Human Approach Big Data and AI

Data no longer just capture politics, economy, culture and lives. Today data is extended and ingrained in society in increasingly complex digital systems developed to contain and make sense of big amounts of it and to act on that knowledge. These digital data systems form a key component of decision-making in politics, culture, industries, and on life trajectories, and consequently they are also the center of power negotiations between different interests. As such, power is at the center of ethical concerns with data systems.

In this study, I present a “data ethics of power” as an action-oriented analytical framework concerned with making visible the distribution of power and power relations in the Big Data Society and the conditions of their negotiation and distribution, in order to point to design, business, policy, social and cultural processes that support a human-centric distribution of power. With point of departure in my immersion in ‘data’ and ‘Artificial Intelligence’ policy, industry and civil society activities of the 2010s in Europe, which among others included the EU’s High-Level Group on AI and the Danish government’s first data ethics expert group, I investigate the role of data ethics and a human (-centric) approach to the ethical governance of AI and big data.

In the first two parts of the study, I explore the power structures for human agency and experience of what I refer to as respectively Big Data SocioTechnical infrastructures (“BDSTIs”) and Artificial Intelligence SocioTechnical Infrastructures (“AISTIs”). I describe these as not only a critical concern of a ‘data ethics of power’, but of, what I refer to as “data ethical governance” in general. In the last part of the study, I answer two quintessential sub-questions when exploring the role and importance of human ethical agency and responsibility in the Big Data Society: Why is a ‘data ethics of power’ important? How can a ‘data ethics of power’ achieve the ‘good society’? The answers to these two questions constitute the formative framework for a ‘data ethics of power’.

The analytical investigation of the emerging cultural and social coalitions of the 2010s concerning data ethics, their negotiation processes and emerging institutionalization in policy, law, governance and data innovation practices, I develop in three research articles: “Making Sense of Data Ethics – The powers behind the data ethics debate in European policymaking”, “Culture by Design - A Data Interest Analysis of the European AI Policy Agenda” and “A Framework for a Data Interest Analysis of Artificial Intelligence”.
Dataetik - en Menneskecentrisk Tilgang til Magt i Big Data og Kunstig Intelligens

Politik, økonomi, kultur og liv udspiller sig i dag i data. Data, som bearbejdes i komplekse digitale systemer udviklet til at lagre og finde mening i store mængder data og til at reagere på denne viden. De digitale datasystemer udgør en nøglekomponent i beslutningsprocesser i samfundet, og derfor er de også centrum for magtforhandlinger mellem forskellige interesser. Magt og magtdynamikker er således kernen af dataetik.

I afhandlingen præsenterer jeg en "Dataetik om Magt", som er en handlingsorienteret analytisk ramme, der beskæftiger sig med at synliggøre fordelingen af magt og magtforhold i big data samfundet samt vilkårene for magtforhandlinger og distribution for at kunne pege på design, virksomheds, politik, sociale og kulturelle processer, der understøtter en menneskecentrisk fordeling af magt.

Med udgangspunkt i min deltagelse i 'data’ og 'kunstig intelligens’ politiske, industrielle og civilsamfundsaktiviteter i 2010’erne i Europa, som blandt andet omfattede EU’s High-Level Expert Group on AI og den danske regeringers første dataetiske ekspertgruppe, undersøger jeg den rolle som dataetik og en menneskecentrisk tilgang kan og burde have i udviklingen af big data og kunstig intelligens.

I de to første dele af undersøgelsen udforsker jeg magtstrukturer for menneskers handlekraft med det, jeg beskriver som henholdsvis Big Data SocioTekniske Infrastrukturer ("BDSTI") og Kunstigt Intelligente SocioTekniske Infrastrukturer ("AISTI"). Jeg beskriver disse som omdrejningspunktet for ikke kun en 'Dataetik om Magt’, men for dataetisk 'governance’ generelt. I den sidste del af undersøgelsen besvarer jeg to delspørgsmål: Hvorfor er en 'Dataetik om Magt’ vigtig? Hvordan kan en 'Dataetik om Magt' opnå det ‘gode’ samfund? Svarene på disse to spørgsmål udgør den formative ramme for min 'Dataetik om Magt’.

Den analytiske undersøgelse af 2010’ernes nye kulturelle og sociale koalitioner vedrørende dataetik, deres forhandlingsprocesser og nye institutionaliseringer inden for politik, lovgivning, regeringsførelse og datainnovationspraksis udvikler jeg i tre forskningsartikler: "Making Sense of Data Ethics – The powers behind the data ethics debate in European policymaking”, "Culture by Design - A Data Interest Analysis of the European AI Policy Agenda” and “A Framework for a Data Interest Analysis of Artificial Intelligence".
# Table of Contents

**INTRODUCTION** ........................................................................................................... 6
  Arguments ......................................................................................................................... 9
  Structure .......................................................................................................................... 13
  Articles ............................................................................................................................ 15
**TERMINOLOGY** ............................................................................................................ 17

**CHAPTER 1 IMMERSION AND METHODOLOGICAL REFLECTIONS** ............. 22

  EMBEDDED RESEARCH AND IMMERSION ......................................................... 23
    1.1 Methodological and Conceptual Framework of the Study ............................................. 27
    1.2 Immersion, Dwelling and Moral Action ........................................................................... 28
    1.3 Multi-Level Analysis .................................................................................................. 32

  RESEARCH QUESTIONS, PARTICIPATIONS AND INTERACTIONS .................. 35
    1.4 Participation, Stakeholders and Interests: A Self-Reflexive Account ......................... 36
    1.5 Interactions with Other Participants in the Research Field ........................................... 41

**PART I POWER & BIG DATA** ..................................................................................... 44

  ABOUT PART I ............................................................................................................... 45

**CHAPTER 2 BIG DATA SOCIOTECHNICAL INFRASTRUCTURES (BDSTIS)** .... 47

  BIG DATA SOCIETY ....................................................................................................... 50
    2.1 The Imagination and Politics of Space and Infrastructure ............................................. 51
    2.2 The Narrative of a European Digital Infrastructure ...................................................... 55
    2.3 The Power Infrastructure of the Information Society .................................................. 58

  ETHICS IN THE SURVEILLANCE SOCIETY ......................................................... 63
    2.4 Liquid Surveillance ..................................................................................................... 63
    2.5 Big Data Ethics .......................................................................................................... 66

**CHAPTER 3 ETHICAL GOVERNANCE AND SOCIOTECHNICAL CHANGE** .... 69

  SOCIOTECHNICAL CHANGE .................................................................................... 72
    3.1 Scientific Knowledge and Practice in Paradigm Shifts ............................................... 72
    3.2 The Complexity of Change: Conflicts, Negotiation and Compromise ....................... 74
    3.3 The Four Phases of Sociotechnical Change .................................................................. 75
    3.4 Ethics in the Policy Vacuums of Sociotechnical Change ............................................. 79

  GOVERNANCE ............................................................................................................... 81
    3.5 Internet Governance .................................................................................................. 81
    3.6 Ethical Governance ................................................................................................... 87
    3.7 Data Ethical Governance ............................................................................................ 89

**PART II POWER & AI** .............................................................................................. 95

  ABOUT PART II ............................................................................................................. 96

**CHAPTER 4 ARTIFICIAL INTELLIGENCE SOCIOTECHNICAL INFRASTRUCTURES (AISTIS)** .......................................................... 98
CHAPTER 5 DATA INTERESTS AND DATA CULTURES

INTERESTS AND TECHNOLOGY
5.1 Stakeholder Interests
5.2 Interests in Social Contexts
5.3 Interests and VSD
5.4 Data Interests

CULTURE AND TECHNOLOGY
5.5 Cultural Values and Technological Style
5.6 Culture and Power
5.7 Data Cultures

PART III HUMAN POWER & DATA ETHICS

ABOUT PART III

CHAPTER 6 WHAT IS DATA ETHICS?

COMPONENT 1: WHY IS A DATA ETHICS OF POWER IMPORTANT?
6.1 The Human(-Centric) Approach
6.2 Henri Bergson’s Human Approach
6.3 Henri Bergson’s Process Ontology
6.4 Intuition as Method
6.5 Data Ethics as a Whole Way of Life and Culture
6.6 The Open Society

COMPONENT 2: HOW CAN A DATA ETHICS OF POWER ACHIEVE THE ‘GOOD SOCIETY’?
6.7 Data Ethics of Power as Practice

CONCLUSION
The Study’s Three Key Contributions
The Structure of the Study
INTRODUCTION

“Imagine a piece of music which expresses love. It is not love for any particular person. Another piece of music will express another love. Here we have two distinct emotional atmospheres, two different fragrances, and in both cases the quality of love will depend upon its essence and not upon its object. Nevertheless, it is hard to conceive a love which is, so to speak, at work, and yet applies to nothing. “

(Bergson, 1932/1977, p. 254-55)

WHY IS DATA ETHICS considered more urgent now than it was ever before? One might say that there is nothing extraordinary about data (or ethics for that matter). The representation and control of life and being in material and symbolic form are only human. Certainly, there is nothing new about data systems reinforcing power dynamics in society or even creating new power relations; furthermore, this relation between humans and their data has always had social and ethical implications. Yet, the transformation of all things into data as an effortless, costless and seamless extra layer of life and society is a recent and new development, which I in this study argue also requires a certain type of ethics approach and awareness from us. Data at the time of writing no longer just captures politics, the economy, culture, and lives - data are an extension of these. Data are ingrained in society in multiple forms in increasingly complex digital systems, which have been developed to contain and make sense of large amounts of data and to act on that knowledge. These digital data systems form a key component of decision-making in politics, culture, and industries, and also on life trajectories, and consequently, they are also the centre of power negotiations between different interests. As such, power should be at the core of any ethical concern with data systems.

I have been immersed in the internet governance and digital rights policy, industry and civil society communities and debates for close to two decades now. Today, in late 2020, the conversation on data ethics and online privacy has matured from a small nuisance and pebble in the shoe of big data actors
and enthusiasts into a main topic of public debate and a crucial policy agenda item in Europe and beyond. Remarkably, an ethical stance on data and data innovation has transformed into what is considered Europe’s competitive position on a global geopolitical stage.

At present, a range of societal stakeholders from industry, academia, and civil society to governments and intergovernmental organisations are presenting the various critical problems in our current data infrastructures: data bias and a lack of multicultural representation, the disempowerment of consumers, democracy challenged by black box algorithms, flawed IT security and data protection, data monopolisation, voter manipulation and many more. Furthermore, the lost opportunities of locked-in data for science, public service, business and in particular the technical development of AI systems are presented as major problems in the state of affairs. Moreover, for each problem raised, an often highly specific solution to a specific problem is suggested. Interests are of course bound to clash, because not all problems with our data reality converge and neither do their solutions, and this is usually the moment when real controversy arises, when power takes material form. It is in these moments of negotiation and controversy that compromises are made and actual sociotechnical transformations emerge.

At times it seems too overwhelming. The complexity of issues, challenges and the compromises we have to make. However, not all is lost to the state of affairs. This study is built on the idea that humans are in control of these compromises—but only if we take into account the entire field of powers and interests invested in a sociotechnical transformation, including our own. Back in the early 2000s when I worked with children’s online empowerment and protection, the debates and negotiations were – in very similar ways – taking the shape of different problems and solutions, formulated and proposed within various constellations of powers and interests. Time and again I also experienced value and interest conflicts, compromises and ethical implications of one-dimensional governance approaches that did not manage to encompass the complexity of sociotechnical transformations and their equally complex ethical and social implications. I saw and assessed countless technical solutions, such as parental filters, age verification for social media, and online ‘walled gardens’; legal and policy solutions, such as child-specific legal requirements, governmentally and European Union (EU)-induced industry self-regulatory activities and dedicated investment
schemes; age-appropriate labelling; educational initiatives, such as the inclusion of digital competences in national school curricula and online campaigns (“You are what you upload” and “You set the limit”); and of course new collaborations and power constellations between, for example, states and digital tech industries.

Accordingly, when in 2014 I left my job at the Danish Media Council for Children and Young People to take on a more general approach to the development and governance of big data, I tweeted, tongue-in-cheek:

“T’ve invented a child online protection button. Now all kids are safe and happy and I’m out of a job”.

Of course, nothing was further from the truth. Many critical actions were taken in terms of mitigating risks and empowering children in a new digital era. Numerous problematic actions were also taken with far-reaching ethical and social implications, such as the development of AI systems for content moderation that intensify the surveillance powers of industries and states, child online protection laws that do the same, technocentric and privacy-invasive digital adoption and educational initiatives in schools driven by tech giants.

However, in all of this I think the greatest challenge we had back then, and what I also see today in the more general ‘internet governance’ efforts of policy and decisionmakers in state, intergovernmental, technical, industry and civil society communities, is that when attempting to orchestrate the symphony of sociotechnical acts and agents that comprise sociotechnical change, we have repeatedly failed to address its complexity. Our vision is clouded by our particular problems and brilliant solutions, and therefore, we do not see the core problem of sociotechnical change, which is the very shape of power dynamics that effect change and outline the direction of sociotechnical development. We do it over and over again. We undervalue the complexity of the powers embedded in our sociotechnical environments and the multifarious ethical and social implications. We neglect to look beyond our own interest, fields of application and expertise, and because of this we create new problems and implications for ourselves, and even produce new ethical problems. Time after time, we overvalue our own – although predominantly well-meaning, but also often very limited – perspectives on problems and solutions. We fail to coordinate, to translate amongst each other and most crucially to assess broader social and ethical implications.
With this study, I have attempted to find a common ground for the debates and negotiations that were taking place in the early 21st Century regarding the development and status of big data and AI sociotechnical environments by spelling out a “human approach” (or what is often referred to in policy and advocacy discourse on AI and big data ethics as a “human-centric/centered approach”), which I refer to as a “data ethics of power”. I guess that is what an applied data ethics essentially is – or at least should be – all about. To find ways to build things, act and govern in a holistic manner that benefits human societies, or what I in this study, with reference to the French philosopher and vitalist Henri Bergson, seek to advocate as open love. Love in the most literal and figurative sense of the word. Love for humanity. Love for the planet and its inhabitants. Love for each other. Love for the universe and beyond. Love without a specific interest. Just love. Nothing more and nothing less.

Arguments

In this study I present a formative framework for a ‘data ethics of power’ (Hasselbalch, 2019). I investigate digital data systems as sociotechnical infrastructures that reinforce and distribute power. As such, they are today also the locus of different powerful interests—corporate, governmental and even scientific. In all of this, the human being’s interest has a very low priority, and at times it has none. This needs to change. With a ‘data ethics of power’, power is not just an arbitrary condition, but rather the material we act upon. A ‘data ethics of power’ is a direct response to our human experience of very present and immediate power dynamics embedded and distributed in sociotechnical big data systems and the conditions and practices that shape them. The aim is to renegotiate asymmetries of power.

But what does this actually mean? Imagine an AI robot that sieves through pictures containing predominantly white faces deciding what human ‘beauty’ means. Ponder on an online search system that learns from news articles to recognize words such as ‘nanny’ and ‘receptionist’ as female and words such as ‘architect’ and ‘financier’ as male (Bolukbasi et al., 2016). Consider an AI assessment model that scores an otherwise brilliant teacher badly because it cannot interpret the social and human dimension of the teacher’s work. Reflect on a mass surveillance global intelligence network or a mass political profiling campaign enabled invisibly by social networking
sites with the personal data of millions of people. These sociotechnical data systems and practices are ethically questionable. Most often they are ‘unfair’, certainly they are not morally ‘good’ but ‘bad’, and in some contexts they might even be deemed illegal. However, data ethics is not a legal assessment, neither is it a moral evaluation of the good or the bad. A ‘data ethics of power’ is concerned with making visible the power relations embedded in big data and AI sociotechnical infrastructures.

I investigate the role of data ethics, technology, culture and society in shaping power structures for human agency and experience and delineate a ‘data ethics of power’ as an essentially ‘human approach’. The ‘human approach’ is first and foremost a critical reflection of the power of technological progress, of the big data and AI sociotechnical systems we build and imagine. It addresses the role of the human as an ethical being with a corresponding ethical responsibility for not only the human living being but also for life and being in general. The ‘human approach’, therefore, also encourages the human interest in the data of big data and AI systems through the involvement of human actors in their very data design, use and implementation.

That is, I argue that data ethics has to become more than just a moral obligation, a set of programmed rules—it has to be human. We can formulate data ethics guidelines, principles and strategies, and we can even program artificial agents to act according to their rules. However, to ensure a human-centric distribution of power, data ethics must take the form of culture, to become a cultural process, lived and practiced as a way of being in the world. As such, a ‘data ethics of power’ first and foremost addresses the cultural conditions and structures of power, rather than the sole value properties of technology design.

With the point of departure in my immersion in the data and Artificial Intelligence policy, industry and civil society activities of the late 2010s in Europe, which among others included the EU’s High-Level Group on AI (HLEG) and the Danish government’s first data ethics expert group, I investigate the role of data ethics and the human-centric/centered approach in the governance of AI and big data. The core research questions for this investigation were sparked by my curiosity in the context of my work in this field: What is data ethics? How can human governance be informed by data ethics? Why was data ethics explicitly included in internet/information society governance and policy in Europe in the late 2010s?
I make a case for an interdisciplinary data ethics. In the late 2010s, ‘data ethics’ was predominantly theorized as a branch of philosophy evaluating moral problems of data and algorithms and articulating morally suitable solutions. In this study, I do not intend to outline an exclusive theory, but rather acknowledge the valuable contributions from a range of fields of studies regarding the ethical implications of a Big Data Society. These fields include applied ethics, legal studies, cultural studies, critical data studies, information, and social and political sciences. Importantly, my interdisciplinary approach to data ethics also encompasses a societal movement in policy, civil society organisations and technology development as well as their applied data ethics practices, such as networks, expert groups, products, technologies, surveys and tools. Crucially, I argue that although seemingly incoherently dispersed in subtopics, fields and practice, these data ethics theories, methodologies, analyses and practices can in fact be delineated in a shared ‘data ethics of power’ conceptual framework.

The key theoretical framework of this study springs from studies that treat the power (moral or political properties; Harvey, 1990; Winner, 1980; etc.) of technologies as dynamic concepts in constant negotiation with human, societal and cultural factors (Bijker et al., 1987; Misa, 1988, 1992, 2009; Bijker & Law, 1992; Hall, 1997; Edwards, 2002; Castells, 2010; Harvey et al., 2017).

The understanding is that while technology may impose on humans and human societies, humans have also time and again imposed on technology, which they may choose to do with intention and direction. Humans create laws, policies and standards; we educate and programme, hack and revolt. This is an important view on technological development and change as it empowers human governance efforts and prevents a single-sided analysis of the components of the sociotechnical development we target with governance initiatives. Combined with a cultural studies approach, I here consider culture one such component of in particular “ethical governance” (Rainey & Goujon, 2011; Winfield & Jirotka, 2018). That is, culture as shared systems of meaning making, as “systems of representation” or “conceptual maps” (Hall, 1997) of power and interests that compete in shaping a “technological momentum” (Hughes, 1983, 1987). Specifically, I investigate the competing data cultures that in the late 2010s were framing our various practices with data as a crucial focus of what I call “data ethical governance”. 
I apply a “multi-level analysis” (Edwards, 2002; Misa, 2009). While understanding big data and AI sociotechnical systems as particular and historically situated, I simultaneously consider their development continuous and open-ended, that is, as components of society and human history at large. I therefore move between different micro, meso and macro plateaus in an effort to simultaneously encompass the embedded power of data design and technology as well as larger social, political and cultural dynamics and conditions.

Accordingly, I consider data ethics in two ways: as, what I refer to as respectively “spaces of negotiation” and “critical cultural moments”. Data ethics spaces of negotiation have a material presence in society. Their role is to enable critique and negotiation. They are possible when systems (material/immaterial and technological/cultural) clash and controversy arises. Data ethics critical cultural moments have special human characteristics. They are possible when human memory and intuition are privileged and provide time and space to tinker. These data ethics spaces and moments can be identified on different scales of time and space, which are intrinsically intertwined. The ability to encompass the entire movement between these different scales of time and space is the essence of the holistic approach of ‘data ethical governance’.

I must admit that this study’s investigation is limited by my Western/European perspective. I speak from a privileged socioeconomic position in a global environment of societies that move at different rhythms and are positioned with different advantages or disadvantages in the evolving big data sociotechnical systems of our age. Questions of power distribution, domination and advantages through big data sociotechnical systems are core components of a ‘data ethics of power’. However, my own analysis fails to represent the essential experiences of the people, communities and cultures that are most disadvantaged by the global big data structures of power. I examine the specific ethical implications and problems of Western (welfare) societies, which are quickly adopting and evolving with big data advanced sociotechnical infrastructures. My analysis, therefore, does not in any way represent the ethical implications and problems specific to many other societies and cultures worldwide, and I thus also neglect the formulation of governance and solutions based on these different situated experiences. This is a limitation of this study that I hope future ‘data ethics of power’ studies based on different cultural and
socioeconomic experiences can amend.

Finally, I have a more informal plea. Data ethics is not only about power—it also is power. We might say that data ethics is what everyone knows but at the same time no one can really say is theirs. However, it is exactly because data ethics does not want to belong, because it resists ready-made concepts, that it becomes everyone's declaration. Governments profess over the good and effective smart Big Data Society; individuals avow to an online life construed by endless streams of personal data; industries and scientists state the value of big data; and so on. All have an interest in data and consequently also in data ethics. Indeed, data ethics is power—power to point out the problems and their solutions, to set the priorities for how we want data technologies in our human lives and in society. I ask, what if we decide to set data ethics free—to uproot the very conceptualisation of the term as the moral obligation of someone or something to solve a specific problem? That is, in a way such that data ethics truly becomes a method and practice for humans to critically challenge the power embedded in data technologies, their set priorities and restraints, and to find different problems and new solutions in the very conditions of the big data reality we live in. This is what I attempt to answer in this study.

Structure

In this study, I present three core competing structures of power, each with their shape and style each with their data cultures. I have therefore structured the main body of work in three parts, with each an emphasis on the three different characters of power:

1. Power & Big Data
2. Power & AI
3. Human Power & Data Ethics

In part I of the study, I address the first kind of power, big data sociotechnical infrastructures (BDSTIs) and their cultures and environments. BDSTIs are constituted physically with fibre cables that run across the globe, enabling data collection and access across geographic territories and jurisdictions, and virtually in spaces of flows around which dominant societal functions are increasingly organised (Castells, 2010). BDSTIs also constitute a
redistribution of power facilitated by these new technologically mediated configurations of space and time. To design and shape the infrastructural components of BDSTIs is here an essential form of power, and for example surveillance powers of state and industry actors are embedded in BDSTIs as a key property of their architecture and design (Haggerty & Ericson, 2000; Lyon, 2001, 2010, 2014, 2018; Hayes, 2012; Bauman & Lyon, 2013; Galic et al., 2017, Clarke, 2018).

In the second part of the study, I focus on the type of power that is concentrated in the emergence of big data artificial intelligence sociotechnical infrastructures (AISTIs). These are first and foremost evolutions of the analytical capabilities of BDSTIs, and constituted as BDSTIs, but with components designed to sense the environment in real time, learning and evolving with autonomous or semiautonomous agency.

The two types of sociotechnical infrastructures described in parts 1 and 2 of the study constitute two forms of power that work in different dimensions of human reality and society. While BDSTIs primarily act in our space by transforming all into immobilised digital data, AISTIs also occupy our time by acting on that data to actively shape the past and present in the image of the future. Accordingly, I propose that a core concern of a ‘data ethics of power’ should be with AISTIs’ and BDSTIs’ constitution as cultural systems of a type of social ordering, in which interests of dominant actors in society are spatialised and immobilised and thus more difficult to criticise and renegotiate.

Finally, in the last part of the study I address human power, which I claim is at the core of a data ethics of power. The human approach of a ‘data ethics of power’ concerns the role of the human as an ethical being with a corresponding ethical responsibility. However, ethics requires spatial and temporal conditions to flourish. Human power is, in the face of the 2010s, sociotechnical changes in constant negotiation with the power of BDSTIs and AISTIs. This therefore implies an applied data ethics in design and governance that ensures the involvement of human actors in the very data design, governance, use and implementation of sociotechnical data systems.
Articles

Three research articles form the analytical basis of this study. While each article takes on a stand-alone argument and analysis, I consider the three parts of the study on Power & Big Data, Power & AI, and Human Power & Data Ethics a formative framework for the articles’ general interpretation and positioning within a data ethics analysis that moves between macro, meso and micro levels of reflection. I therefore suggest that the three parts are read before the articles.

The three articles can be summarized as follows:

In the first article, titled Making Sense of Data Ethics – The powers behind the data ethics debate in European policymaking (2019), I examine the data ethics public policy initiatives that took shape in the aftermath of the European General Data Protection Regulation (GDPR) reform. Addressing a general uncertainty as to these policy initiatives’ role, I propose that the initiatives are acknowledged as open-ended spaces of negotiation extended over time responding to governance challenges embedded in complex sociotechnical infrastructures. I argue that by examining these data ethics policy initiatives as components of a general infrastructural development’s rhythm, rather than carved in ethical solutions, we create an opportunity for a more transparent value- and interest-aware negotiation of ethical action in the big data era. It is also in this article that I develop an initial framework for a ‘data ethics of power’ as an action-oriented analytical framework concerned with making visible the power relations embedded in the Big Data Society and the conditions of their negotiation and distribution, in order to point to design, business, policy, social and cultural processes that support a human-centric distribution of power.

In the second article of this study, titled, Culture by Design: A Data Interest Analysis of the European AI Policy Agenda (2020), I examine the concept of culture in the context of sociotechnical development with reference to an institutionally framed European cultural interest in shaping the global AI momentum of the late 2010s. In that period, the development of big data sociotechnical systems with AI capabilities were characterised by cultural positioning and global competition invested with interests from across societal groups, regions and sectors. Hence, a political agenda emerged in the EU with an emphasis on a shared European cultural framework for the
development of AI and the data pools upon which the technology strove with an emphasis on ‘ethical technologies’ and ‘trustworthy AI’. In the article, I describe how this cultural interest evolved in a process of public events and the work of the HLEG on AI established by the European Commission.

In the third article, titled A Framework for a Data Interest Analysis of Artificial Intelligence (forthcoming, FirstMonday), I describe AI systems as moral agents in which humans delegate the enforcement of the agency of different interests in society. I argue that recognising this type of delegated agency of interests allows humans to develop and design AI in accordance with the human interest and to make suitable choices. Thus, I introduce the term “data interests” and make a case for a “data interest analysis” of the complex data processing systems of AI, exploring how different interests in data are empowered or disempowered by design. I here use the EU HLEG on AI’s Ethics Guidelines for Trustworthy AI constructively as an applied ethics approach to data interests, and suggest questions that will help resolve conflicts between data interests in AI design within a human-centric data ethical governance framework.
TERMINOLOGY

The following is a delineation of how I use some key terms in this study. The terms are also delineated in more detail throughout the study.

Power
The ‘data ethics of power’ that I formulate in this study is concerned with a redistribution of power facilitated within new technologically mediated configurations of space and time (or what I refer to as “BDSTIs” and “AISTIs”). It recognizes value and interest cultural power struggles and negotiations as a core component of sociotechnical change and governance (Hughes, 1983, 1987).

This conceptualisation of power of the study stems most profoundly from surveillance and critical data studies that address the state of power in the Big Data Society on the level of the micro design of systems, business, state and engineering data practices (Bowker & Star, 2000; O’Neil, 2016; Angwin et al., 2016; Hasselbalch & Tranberg, 2016; Eubanks, 2018; Noble, 2018 etc.) and on the level of macro societal sociotechnical change (Lyon, 2001, 2010, 2014, 2018; Bauman & Lyon, 2013; Castells, 2010, etc.). In particular, I address power in terms of its “liquidity” (Baumann, 2000; Bauman & Haugard, 2008; Lyon, 2010; Bauman & Lyon, 2013, Castells, 2010). That is, I am concerned with a power that is concentrated and engineered by a few power actors (Zuboff, 2019 etc.), yet also increasingly self-sustained, re-engineered and evolving in (surveillance) cultures (Lyon, 2018) of use, design, governance and imagination, and therefore difficult – but not impossible – to change.

The study is also based on a more general conceptualisation of power that I derive from a tradition within cultural studies that addresses the uneven distribution of power in cultural representation, cultural practices and products (Williams, 1993/1958; Harraway, 1985/2016; Agger, 1992; Gilroy, 1996/2012, 1987/2012; Hall, 1980, 1994, 1997 etc.). Crucially, in this perspective, cultural power is never stable and can always be challenged and redistributed (Williams, 1958/1993).

Sociotechnical
Technology is always part of society, just like society is always part of
technology. This also means that one cannot understand one without the other. Technology is not only design and material appearance but also sociotechnical; that is, a complex process constituted by diverse social, political, economic, cultural and technological factors (Hughes, 1987, 1983; Bijker et al., 1987; Misa, 1988, 1992, 2009; Bijker & Law, 1992; Edwards, 2002; Harvey et al., 2017).

**(Sociotechnical) Infrastructures**
Infrastructures are the virtual and material sociotechnical organisations of the space of societies. They are engineered and directed, but they also evolve in social, economic, political and historical nonengineered dynamic contexts. Specifically, a sociotechnical infrastructure is a particular type of human-made space which is the material and immaterial, engineered and nonengineered processes that evolve in a space of negotiation and struggles between different societal interests, imaginations and aspirations (Star & Bowker, 2006; Bowker et al., 2010; Harvey et al., 2017).

**BDSTIs (Big Data SocioTechnical Infrastructures)**
In part I of this study, I introduce the term “BDSTIs” to refer to sociotechnical infrastructures based on big data technologies. They are the primary infrastructures of the flows (Castells, 2010) of global economies and societies cutting across geographic territories, legal jurisdictions and cultures. In the 2010s, BDSTIs were increasingly representing and constituting global societies and environments as the mundane background against which social practice, social networking, identity construction, economy, culture and politics were conducted. They were in part institutionalised, in systems requirements standards for information technology (IT) practices, and in regulatory frameworks for data protection, and they were invested with human imagination about the challenges and opportunities of big data.

**AISTIs (Big Data Artificial Intelligence SocioTechnical Infrastructures)**
“AISTIs” is a term I use to describe an evolution of the analytical capabilities of BDSTIs. AISTIs are constituted as BDSTIs but with components designed to sense the environment in real time, learning and evolving with autonomous or semiautonomous agency. While BDSTIs act in space by transforming all into immobilised digital data, AISTIs also occupy time by
acting on that data to actively shape the past and present in the image of the future. In part II of the study I in particular focus on the history, ethics and development of AISTIs.

Culture
There are two sides to culture: (1) It is a system that brings together communities with shared conceptual frameworks and resources, and it is an active system with specific priorities, goals and ways of organising the world that are actively imposed in society. (2) Culture is “a whole way of life” (Williams, 1993/1958). It consists of prescribed dominant meanings, but importantly culture is also the negotiations of these meanings. That is, culture is not just one; it is multifaceted, institutionalised and formalised, and practiced by dominant groups in society, as well as subcultural and practiced by, for example, minority groups in society. Thus, culture is never stable; it is from the outset a constructed system of meaning making and is therefore also always up for contestation and social negotiation (Williams, 1993/1958; Gilroy, 1996/2012; Hall, 1997).

Data Cultures
I describe the cultures that frame data science, practice and governance as “data cultures”. They are culturally coded conceptual maps of the engineers, data scientists and designers of data systems; deployers of data systems; legislators of data systems; and users of data systems. They are not always shared and they may even be in conflict (Hughes, 1983, 1987; Friedman & Hendry, 2019). Data cultures are interrelated with societal power negotiation and struggle. The very practices of data scientists and designers are, for example, framed within specific informal or institutionalised cultural systems of meaning making. Accordingly, the very practice of developing a data system and design is a cultural practice (Bowker, 2000; Bowker & Star, 2000).

The Human (-Centric) Approach
The human-centric or human-centered approach was a popular term in the late 2010s’ policy and advocacy discourses on the ethics of AI and big data, used as a way to recenter the socio-technical developments in these fields on the human interest. In this study, I further explore and conceptualise this term, but I refer to it as the “human approach”. I do this to emphasise the
role of the human as an ethical being with a corresponding ethical responsibility for not only the human living being but also for life and being in general.

In practical terms the human (centric/centered) approach is associated with the human interest in the data of AI through the involvement of human actors in the very data design, use and implementation of AI.

The human approach of a data ethics of power in specific constitutes a critical reflection on the power of technological progress as well as the big data and AI sociotechnical systems we build and imagine.

**Values and (Data) Interests**

Values are “idealized qualities or conditions in the world that people find good” (Brey, 2010, p. 46). They are represented in power struggles over different cultures and worldviews. Interests are held by different actors and represent social power struggles over material things such as resources. A data ethics of power is particularly concerned with the interests in data invested in data design and governance (“data interests”, article 3 of this study). Values and interests are core components of sociotechnical change. (Hughes, 1983, 1987; Friedman & Nissenbaum, 1995, 1996, 1997; Friedman, 1996; Shapiro, 2005; Friedman et al. 2006; Flanagan et al. 2008; Spillman & Strand, 2013; Friedman & Hendry, 2019 etc.).

**Moral Agent/Agency, Ethical Agents/Agency**

The concept moral agency is often used interchangeably with ethical agency. However, in this study I make a distinction between the two to emphasise the difference between two different capacities. I understand a moral agent as one that can only enforce and act according to moral prescription and determination. For example, ‘intelligent’ nonhuman agents (AI agents) are moral agents, but they are not ethical beings. This is also why I consider data ethics a human responsibility only. (Bergson, 1932/1977; Searle, 1980, 1997; Smith, 2019; Amoore, 2020; Pasquale, 2020).

It must be noted here that when referencing other authors, I must at times use the terms moral and ethical interchangeably. In these cases, my distinction will be clear from the content.

**Human actors/agents and nonhuman actors/agents**

I deliberately make a distinction between human and nonhuman agents and
actors. However, my intention with this very rough distinction between the human and nonhuman is not grounded in a technological nor cultural determinism; rather, it is a semantic trick aimed at disclosing the limits of the moral agency of AISTIs in particular and, in this connection, the importance of human ethical agency and power to change and govern sociotechnical development. Accordingly, despite my distinction between the two, I do recognize technological artifacts as extensions of human agency and intent as well as their increasing indistinguishability with human environments.

**Ethical/Data Ethical Governance**

“Ethical governance” (Rainey & Goujon, 2011; Winfield and Jirotka, 2018) is a multi-actor, reflexive, open-ended (Hoffman et al., 2017, Harvey et al, 2017) and agile process designed to ensure the “highest standards of behaviour” (Winfield and Jirotka, 2018). It goes beyond just good and effective governance. I define “data ethical governance” as a form of ethical governance that in specific addresses the complexity of the Big Data Society with infrastructural practices that create human-centric data cultures.
CHAPTER 1
Immersion and Methodological Reflections
EMBEDDED RESEARCH AND IMMERSION

This study takes its point of departure in my immersion in the internet governance and digital rights policy, industry and civil society communities. Some introductory reflections and a history of these years of experience are therefore crucial as they provide essential insight and context to my position as an embedded researcher, as well as to the method and approaches I used in my research.

From 2005 to the end of 2014, I worked in an Awareness Centre on children and young people’s use of new online technologies in the Danish Media Council, which is part of the Danish Film Institute, a governmental institution under the Danish Ministry of Culture. When in the early 2000s I started working with these issues, the average adult population mainly knew the internet as a communications platform for emails, basic search and news. However, children and young people were quickly adopting the first emerging online social networking sites, and this early digital extension of everyday life and society was accordingly by the average adult population first and foremost perceived as an inaccessible, incomprehensible and secret world of youth. Therefore, the Media Council, whose primary function was to implement film classification in Denmark, had in late 2004 established the Awareness Centre to address the digital evolution of children’s media environments.

The Awareness Centre was part of the pan-European network of safer internet centres, Insafe, co-funded by the European Commission’s Safer Internet Programme (later changed to the Safer Internet Plus Programme and then Better Internet for Kids). These programs were created to address the risks posed by the internet to children and young people, and also to provide tools, knowledge and skills to empower children, their parents and teachers to make the most of its opportunities. This was a new and rapidly evolving public policy area and the work – by the human and nonhuman actors shaping it – equally so.

Generally, it became increasingly evident that the internet was not just technically evolving in a linear direction as a bottom-up engineering or
citizen project or a top-down governmental governance initiative. It was steered and nonsteered, engineered and nonengineered, accepted and revolted against, autonomous and controlled—all at the same time. Thus, to ensure opportunities while mitigating the risks, or in other words to attempt to govern this sociotechnical digital development, was not a simple process.

In the 10 years I worked with the Awareness Centre, I was on the ground with teachers, parents and children and in meeting rooms and conference halls cooperating with civil society, governmental and industry stakeholders. To understand the secret online world of youth, we conducted studies. Moreover, to help adapt national and international policy, we participated in multistakeholder initiatives and events advocating youth empowerment and human rights such as the UN Internet Governance Forum. To change institutional frameworks, we worked with governments and institutions to advocate, for example, for the inclusion of digital competences and ‘web ethics’ in school curricula. To provide concrete tools and to empower children, parents and teachers, we conducted crypto parties and workshops with parents, developed educational materials, and created global awareness campaigns.

These were also the early years of the popularisation of the privacy movement, which had existed since the beginning of the internet’s history in the 1990s in more technical activist communities. Privacy was now considered a form of power that citizens should demand by using alternative services, including early privacy enhancing technologies to protect themselves against state and commercial tracking and surveillance. However, at one point I realised that empowering and informing the users of the internet was not enough. I became particularly concerned with the way in which the initial freedom away from adults’ prying eyes, which youth had been experiencing with the introduction of the internet, gradually had transformed into a new form of control by powerful actors, such as social media tech giants. What really concerned me was their presence everywhere—at our own events and meetings, in public consultations, in policy initiatives. It was as if their business design and model for the evolution of the Age of Big Data was the only formula possible.

I therefore started focusing on alternatives to the very design and business models of these services inspired by the early critical voices in the field. Since the early 2000s, civil society and academic actors in particular had worked to have human rights issues included on the official internet
governance agenda in processes such as the World Summit of the Information Society (WSIS) (2003, 2005) and the UN Internet Governance Fora (IGF) (2006 - ). However, it was not until 2013 that the United Nations General Assembly affirmed that the same rights that people have offline must also be protected online (UN, 2013). Even then, in the more general business and public discourse, human rights online did not take a proper foothold. In fact, privacy was often described as an obstacle to digital innovation, as an old social norm that was preventing an unavoidable digital evolution of society. It was not difficult to understand that the core problem with the implementation of a human right, such as privacy, when implemented in the current sociotechnical infrastructure of the internet, was the very business and technology culture of the internet’s development. In 2014, we therefore established the global “privacy as innovation network” at the UN Internet Governance Forum, bringing together industry, human rights advocates and technology entrepreneurs to explore privacy as an opportunity rather than an obstacle.

In late 2014, I left the Danish Media Council. Together with former journalist Pernille Tranberg, I started exploring a growing movement among technology designers and emerging companies, who were developing and promoting alternative data design and business models based on the preservation of privacy. We understood that one of the core obstacles to privacy was a commercial interest that was shaping the innovation that went into the design of online services. We decided that a way to address this culture was to present a business idea. This idea was that a renewed demand in society, including the new legal requirements for data protection that were being negotiated in Europe, was also a somewhat unexplored market and accordingly behaving ethically with data would be a competitive advantage for companies that were serious about it. Therefore, the book we had set out to write was at first a trend analysis of alternative (to the big tech actors) companies working constructively with online privacy to inspire other companies to do the same.

However, while we were researching the topic and talking to designers, company representatives and policymakers, we understood that privacy was

---

1 See “Privacy is the latest digital media business model” op ed in Politiken, August 2013 https://mediamocracy.org/2013/08/23/data-ethics-the-new-competitive-advantage/
not the only sacrifice society would make if the business development of the internet was not redirected. The social implications were more general in nature. This is when we decided to focus on data ethics as a constructive, holistic approach primarily in business development that went beyond mere compliance with data protection laws; that is, the check list mentality in which social and ethical reflections on the implications of data design was added as a an after thought.

*Data Ethics – The New Competitive Advantage* (Hasselbalch & Tranberg, 2016) was thus first and foremost an action-oriented awareness project targeted at a specific stakeholder group in society. This is also why in 2015, together with two other women, Marianne Steen and Birgitte Kofod Olsen², we created the thinkdotank DataEthics to promote alternative business approaches and provide knowledge, collaboration on data ethics for businesses, educational institutions, organisations, individuals and decision-makers.

At this time, data ethics was still not a layman’s term and the ethical implications of data technology and business were still addressed in public debate – if addressed at all – in terms of privacy implications only. As such, it was still a great struggle for the lone privacy activist to walk into a public debate on social media and the digitalisation of society. Human rights issues of online business were considered an activist topic separate from the debates on big data innovation and disruption that were shaping online business development. However, public discourse was also changing; in particular, the negotiations of the European General Data Protection Reform were increasingly addressed. In late 2015, the European Data Protection Supervisor (led by the late Giovanni Buttarelli) established an external advisory group on the ethical dimensions of data protection. Alongside a growing awareness and attention to the ethics of digital technology, the thinkdotank DataEthics became increasingly involved in the public debate as well as in business and the policy debate.

Thus, when I started this study in 2017, I was already involved in initiatives and participated regularly and actively in this study’s area of interest.

---

²I want to also acknowledge Catrine Søndergaard Byrne who acted as founder and a key member of the managing team of DataEthics.eu from 2017-2019.
1.1 Methodological and Conceptual Framework of the Study

In this section, I combine reflections on the methodological framework for this study, which takes its point of departure in immersion, participation and action in the field of study (Ingold, 2000; Wong, 2009; Bourdieu, 1992, 1997; Goffman, 1974; Atkinson & Hammersley, 1994; Sandiford, 2015; Gray 2013; Tope et al. 2005), with a multi-scalar analysis (Misa 2009, 1988, 1992) that informs the structure of the analyses of the study. In the analysis, I also ground my interpretations in a philosophical Bergsonian framework, which is outlined in more detail in the last chapter of the study.

Embedded research entails the immersion of the researcher in the field of study as an active and engaged participant (Bourdieu, 1977/1997; Bourdieu & Wacquant 1992; Ingold, 2000; Wong, 2009). The approach draws on methodologies characteristic of ethnographic studies, such as participation, holistic data collection, and formal and informal interactions with other participants in the field. Ethnography involves an exploration of cultural phenomena, groups and communities by immersion in the field, drawing on different perspectives to provide a holistic description of contexts and cultural themes (Richards et al., 2012; Paltridge et al., 2016). As such, it relies on different forms of data collection, uses many different sources, and does not study isolated aspects independently of the situation in which they occur (Madden, 2017). The collection of data is not based on “pre-set categories”
but arises out of a general interest in an issue or problem (Madden, 2017). Accordingly, interpretation of data also depends on the very immersion in the field. As such, the researcher is considered a “central research tool” whose position and influence in the field of study and research process must be carefully reflected upon (Madden, 2017, p. 20). The objectivity of the researcher is a much-contested term often correlated directly with the validity of the scientific method. In anthropology and social science (Williamson et al., 1977, p. 90-92), the researcher’s personal biases and the power relations between the researcher and the ‘researched’ are considered core challenges, but are unavoidable dimensions of social research when exploring and uncovering social and cultural phenomena and their dynamics. They create the unavoidable context of an interaction between the perspectives of the observer and the interactions with the observed.

1.2 Immersion, Dwelling and Moral Action

Embedded research can be argued to differ from other ethnographic accounts through its explicit emphasis on the position and immersion of the researcher as a core dimension, and strength of the method, more than a form of weakness that must be overcome (Wong, 2009). In his book The Perception of the Environment (2000), anthropologist Tim Ingold describes the embedded research approach as an “immersion” in an environment. This immersion is, he argues, a way to overcome a distinction between naturalistic and cultural logical research accounts and thus what he considers a false dichotomy between an environment and the individual and social dynamics in which it is experienced.

This means that as researchers we may consider our own situated experiences – and positions – as essential components of our scientific investigations. Or, in other words, the very position of the researcher, which is constituted in relation to other components in the environment, may also be considered a valid component of the research conducted (Sandiford, 2015). As Ingold explains:

“Simply to exist as sentient beings, people must already be situated in a certain environment and committed to the relationships this entails. These relationships, and the sensibilities built up in the course of their unfolding, underwrite our capacities of judgement and skills of discrimination, and
scientists – who are human too – depend on these capacities and skills as much as do the rest of us” (Ingold, 2000, p. 25).

As such, the presumption of an embedded research approach such as Ingold’s is that humans and environments emerge together (Knox, 2017, p. 354). One cannot be in an environment without acting in it. Knowledge emerges as a situated, embodied and engaged practice. In fact, the very immersion can be considered an asset in an action-oriented framework, if accompanied by a particular type of reflexivity of the researcher; that is, reflexivity as a form of “bending back” on herself, not in terms of personal experiences and observations but rather a reflection on the history, conditions and power dynamics of the research conducted (Bordieu & Wacquant, 1992; for an in-depth account of my own reflexive account, see Section 1.4). This view is in essence also an acknowledgement of the moral agency of the researcher. It is a kind of reflexivity that can be described as a constructive force in a political reality in which “…social science cannot be neutral, detached, apolitical…” (Bordieu & Wacquant, 1992, p. 51).

Ingold introduces the concept of “dwelling”, the very immersion of the human being in the environment. Dwelling is a way of being and perceiving an environment that makes us capable of perceiving the environment from everywhere at once (Ingold, 2000, p. 226), or it is to “think movement” (Bergson, 1907/2001, p. 318), to use a Bergsonian term. That is, when dwelling in an environment, we also accept it as “…a continuous itinerary movement” (Ingold, 2000, p. 226) rather than just an immobile instant that we investigate and represent from a detached ‘outside’ position. Ingold here uses intuition as the type of perception or understanding appropriate to access an evolving environment. It is an “intuitive understanding” based on (the researcher) being situated in the environment, informed by a memory that goes beyond the immediate experience: “… based in feeling, consisting in the skills, sensitivities and orientations that have developed through long experience of conducting one’s life in a particular environment” (Ingold, 2000, p. 25). Thus, this type of intuitive perception falls between the rational objective intellect and the subjective experience and skills of the researcher, positioned in an environment, built on memory and an orientation within it. Moreover, we thus may infer from this that the longer the memory, the higher its research value and asset.

In the last Chapter 6 of the study, I delineate a Bergsonian philosophical framework for the study’s ‘data ethics of power’ and ‘human approach’ that
privileges human memory, intuition, reflection and sentience. Specifically, this includes the consideration of a type of philosophical approach and perception, what Gilles Deleuze has also described in his delineation of Bergsonism as “Intuition as Method” (Deleuze, 1966/1991). This is a philosophical approach that challenges stability with the concept of a state of being that is always moving towards an undefined point in the future in a state of becoming, or what Henri Bergson refers to as “to think movement” (Bergson, 1907/2001, p. 318). Intuition is in this context described as a type of perception based on a situated experience with a memory (Bergson, 1896/1991).

Here, I consider a ‘moving’ or action-oriented analysis performed as an analysis of ‘movement’ between different scales of time and experience, and the analysis, essential to the investigation of the complexity of factors that compose a sociotechnical system, as well as the way in which it moves, develops and changes. The core aim of the study is to conduct an action-oriented analysis. This consists of a discussion of sociotechnical change and governance with a delineation of a formative framework for a ‘data ethics of power’ guided by my immersion in the data ethics and AI ethics policy and public debate in the late 2010s.

As described in the Introduction to this chapter, throughout the course of my professional life, I have participated in many different initiatives inside and outside the policy sphere with moving objectives and a range of different stakeholders. These are all integral to my understanding of the history and power dynamics of the field I explore in the present study. However, during the course of the study, which was conducted over a two-year period from late 2017 to late 2019, I participated in the following three initiatives that were particularly formative for my arguments regarding data ethics as spaces of power and interest negotiation and cultural positioning: The IEEE P7006 Standards Working Group on Personal AI Agents.3 The

3 In early 2017, I was approached to help develop the objective (the “PAR”) for a new IEEE standards working group P7006 and to take on the role as vice chair for the group. In 2017, I had been involved in the association’s Global Initiative for Ethically Aligned Design for Autonomous and Intelligent System for a while. Within this initiative a series of standards for incorporating ethical considerations into the design of AI systems referred to the P7000s were developed in multi-stakeholder working groups to provide constructive guidance for engineers. The objective of the P7006 standard established within this initiative we described as one that would address the technical elements required to create and grant access to a personalized Artificial Intelligence (AI) comprising inputs, learning, ethics, rules and values controlled by individuals. I use some of the preliminary results of this group in the second article of this study on the data interests of AI.
Danish Expert Group on Data Ethics, and the EU HLEG on AI. In this study I specifically focus on the HLEG on AI. In 2018 I was appointed to the European Commission’s HLEG on AI, which consisted of 52 experts and stakeholders from academia, civil society and industry. The HLEG’s objective was to support the implementation of a general European Strategy on Artificial Intelligence (2018). This included the development of ethics guidelines for AI as well as policy and investment recommendations. The year after its creation in June 2019, the HLEG delivered two highly influential documents: The Ethics Guidelines for Trustworthy AI and a set of 33 Policy and Investment Recommendations for Trustworthy AI in Europe (I focus on these two deliverables in the second and third articles of the study on the cultures and data interests of AI, respectively).

With its point of departure in my immersion in the internet governance, data ethics and AI ethics public and policy debate of the 2010s, this study represents aspects of a case study in that it is delimited and situated in a particular context (Casanave, 2015). However, my exploration of the topic first and foremost takes its point of departure in a Science and Technology Studies (STS) conceptual framework. Therefore, while understanding sociotechnical systems as particular and historically situated, I simultaneously consider sociotechnical developments continuous and open-ended; that is, as components of society and human history at large. Crucially, within this conceptual framework, I approach technology as simultaneously shaping and shaped by society (Hughes, 1987, 1983; Bijker et al., 1987; Bijker & Law, 1992; Misa, 1988, 1992, 2009; Edwards, 2002; Harvey et al., 2017). This means that the analysis of this study also addresses and moves between different micro and macro plateaus in an effort to simultaneously encompass the embedded power of data design and technology and the larger social, political and cultural dynamics and conditions.

---

4 In March 2018, I was invited to become one of the 12 members of an expert group on data ethics appointed by the Danish government to come up with data ethics recommendations for Denmark. The initiative was part of the government’s strategy for Denmark’s digital growth. The group was set up by the Danish Ministry of Industry, Business and Financial Affairs and a core task of the group was therefore to investigate how Danish companies could make responsible data use a competitive advantage. Later in 2018, the group presented 9 recommendations, which among others included a “data ethics seal” and a “data ethics committee” (Data Ethics Expert Group, 2018). While I do not use my participation in the work of this group, nor the results, explicitly in the study, the power dynamics and interests at play in this period have informed my more general conceptualization of a ‘data ethics of power’. 
1.3 Multi-Level Analysis

The importance of a movement between different levels of analysis to encompass the complexity of sociotechnical change is addressed well by Thomas J. Misa (2009, 1988, 1992). He suggests a “multi-level” analysis that encompasses both micro and macro perspectives, which he does to overcome what he considers a false dichotomy between two different framings: the relation between humans and machines produced by analyses that adhere strictly to either a micro or macro perspective on technology (Ibid. 1988, 2009). Focusing on either the micro dynamics of, for example, designers and engineers of a technology, or on the other hand only on larger macroeconomic or ideological patterns, will present very different and often conflicting views on sociotechnical change, argues Misa. That is, while one will often not see the wider social conditions and power dynamics for change, the other will just as often reduce individual nuances and factors by making sense of them in terms of larger societal dynamics. A multi-scalar analysis conversely acknowledges both perspectives. Misa concludes: “If my intuition is correct that “findings follow framings,” micro-level accounts as well as macro-level accounts can each give us valid if partial insight into the nature of technology and society. Can we take the next step, then, of recognizing these varied scales or levels, analytically moving between them, and combining their perspectives and insights?” (Misa, 2009, p. 367).

Here, Misa draws on Paul N. Edwards’ “multi-scalar analysis” of infrastructure. Edwards (2002) describes this in his discussion of infrastructures as the lived experience of modernity as a method in which we are “...looking at infrastructures simultaneously from a variety of scales of force, time, and social organization” (Edwards, 2002, p. 192). Moving on different scales of time allows us to see the specificities and patterns of sociotechnical change. As Edwards argues, we may detect larger patterns of technological innovation and consolidation on a historical scale, while simultaneously understanding the specific life cycles of the different technologies, such as the railroad and electricity of modernity. In this way, we can simultaneously also understand their political, organisational and cultural contexts.

Three scales of times (micro, meso and macro) are also central to my delineation of a data ethics of power’s ‘spaces of negotiation’ and ‘critical cultural moments’:
On the micro scale, the very design of a technology can be discerned as closed or open to human controversy and negotiation. For example, an AI agent’s algorithms and data processing can be a black box and evolve autonomously without human intervention. Alternatively, it could have a human in the loop design, transparency of design, auditability, and personal data control. Focusing on the design of the technology, I focus on the micro time and space in which it is designed and programmed by human designers or when it is implemented by human users. The micro-scale temporal analysis of a data technology here considers whether the very data design and the design process for this are open for cultural value negotiation. For example: Is the design process locked in an unquestioned technological data culture? Or is it open for critical assessment in terms of the invested interests in the process?

On the meso scale, institutions, companies, governments and intergovernmental organisations will be open or closed to negotiation of the values and cultural frameworks for their practices. Closed entities will move along the natural state of affairs and compliance only with laws. Open entities will establish initiatives and practices dedicated to value negotiation and ethical reflection in addition to legal compliance. Patterns can be identified and when ethical reflection on a specific topic is repeated across initiatives and entities, ‘data ethical governance’ may be introduced.

Lastly, sociotechnical developments and change can be analysed on a macro scale of time. Moments of ethical reflection and negotiation here emerge in between crisis and consolidation of a sociotechnical system in society (Hughes, 1983, 1987; Moor, 1985). These moments are critical as they constitute social negotiation and result in cultural compromises, namely “the technological momentum” (Hughes, 1983, 1987) that a sociotechnical system needs to evolve. They are also crucial to phases of innovation and development as they constitute the transformation of the sociotechnical system that emerges from a quest to solve the critical problems of the system. Hughes refers to “a battle of the systems” in which an old and a new system exists simultaneously in a relationship of “dialectical tension” (1983, p.106-139). He describes this phase as a moment of conflict and resolution not only among engineers but also in politics and law. In these moments of conflict, critical problems are exposed, different interests are negotiated, and they are finally gathered around solutions to direct the evolution of the systems. The new system, or the transformation of the old system, evolves
out of the problems identified and solved in this phase. These critical problems of the systems are not just resolved as technical problems, for example, with the agreement on technical standards with systems requirements, but are in dialogue with political and historical factors.

The multi-scalar analysis informs the structure of this study, moving between these different scales with more attention to the macro and meso scales in part I (Power & Big Data) and on the micro scale in part II (Power & AI). In the last part (Human Power & Data Ethics), I portray data ethics as an ethics concerned with power structures on all three scales. Similarly, while each article of this study takes on a stand-alone argument and analysis, I also consider them as a whole moving between macro, meso and micro levels of analysis. As a result, in the first article I concentrate on public policy data ethics initiatives on a meso scale of time; in the second article I consider the European AI agenda in terms of a cultural positioning in a technological momentum (Hughes, 1983, 1987) of sociotechnical patterns that should be discerned on a macro scale of time; and in the last article I more specifically examine data interests in the micro scale of time of what I refer to in this study as “data designers” of AI.
RESEARCH QUESTIONS, PARTICIPATIONS AND INTERACTIONS

My immersion in and long-term situated experience of policy and public debates on internet governance, data ethics and AI ethics were formative for the delineation of the study, the very research questions I sought to answer, the theory, data collection, the analytical movement between the micro and the macro scales of time, and the action-oriented applied ethics approach. As a point of departure, the ethical problem I set out to solve was based on my own lived experience in the field. Over the course of the years that I participated in the field of interest (before and during the period of study), I experienced a rising awareness of the role of the internet in society, which increasingly included a view on risks and challenges beyond mere technical and functional issues. Notably, during these years a focus by civil society activists and advocates on the privacy risks of data technology, innovation and business transformed into a more generally accepted awareness of data ethics among business and governmental stakeholders as well as the general public. Thus, the study was sparked by my own curiosity in the context of my immersion, action and participation in the field. Why did data ethics catch on in policy and public debates? Why were fundamental rights (such as privacy) not enough? Why were all of us advocating and using the term data ethics in the European public debate and policymaking not satisfied when the GDPR was introduced? Furthermore, when the criticism of the use of the term data ethics emerged, what were the interests? How was the term appropriated? As described in the Introduction, I saw a pattern regarding power: the distribution of power in the data of the sociotechnical systems, and thus the actual data systems and their cultures as a form of power, and not least the interests invested in the very term data ethics as an expression of power dynamics. Accordingly, I delineated three key research questions for the study:

1. What is data ethics?
2. How can human governance be informed by data ethics?
3. Why was data ethics explicitly included in internet/information society governance and policy in Europe in the late 2010s?

To explore these research questions, I investigated initiatives in Europe with explicit reference to digital ethics or data ethics in the period 2015–2019, with a specific focus on the period from 2017 to the end of 2019, which was the period of this study. As a starting point, when data ethics gained traction in public discourse in the slipstream of the implementation of the General Data Protection legislative reform in Europe in the late 2010s, it took at the outset the form of a type of moral assessment of technology design, state and business conduct. Although the power of big technology actors was addressed early by key trend-setting civil society actors, academics and politicians (as I describe in the first article of this study), it was not the central focus of the public debate. At this time, a number of actors and different societal groupings in industry, state and civil society also started proposing core principles for data ethics, presenting data ethics tools, arguments and initiatives (including myself and my own organisation). However, with one foot in the policy field (increasingly invited to take part in policy events and initiatives), another in civil society advocacy and activism and a view into business practices through the focus of my work, I quickly understood that data ethics was more than a moral evaluation and a solution to a specific moral problem. The values (including my own) we wanted to implement in technology practice, and therefore also the ‘data ethics solutions’ (technical and governance), were expressions of our own position in a field of power negotiation.

1.4 Participation, Stakeholders and Interests: A Self-Reflexive Account

In the study, I do not describe, compare, or evaluate individual initiatives that I have participated in. The aim is first and foremost to shift the attention from a moral evaluation of technology to a more holistic governance approach and to provide the theoretical as well as philosophical foundation to do so. I imagine that a skilled ethnographer will be disappointed in me, expecting from this study a detailed exploration of the micro power dynamics of these initiatives. Indeed, this would also have been a valuable analysis of micro power dynamics that could help further the exploration of
how power relations in spaces of group negotiation make out the shape of a data ethics that is anything but a neutral moral evaluation. However, I feared that in doing so I would also have been lost in these micro power dynamics and thus lose sight of the holistic approach and framework that I believe is so urgently required to guide governance and actions in the field.

Nevertheless, here I will meet a few expectations with some reflections on the different power dynamics at play in some of the initiatives I took part in during the period of study to illustrate how power works on many levels—on macro as well as micro and meso scales of time and space.

For example, each of the two policy initiatives I participated in during said period, the Danish data ethics expert group and the HLEG on AI, represented different member selection processes. While there was no application process for the members of the Danish expert data ethics group, and the mode of selection was therefore unclear (even to me who was selected), I had to send a formal application to be selected as a member of the HLEG on AI. Furthermore, the European Commission selected members according to a set of criteria regarding the very composition of the group (individuals appointed in a personal capacity, individuals appointed to represent a common interest shared by stakeholders, and organisations; European Commission J, 9th March, 2018). Yet, in each case, I still do not have a completely clear picture of the relations invested in each of the selection processes. Accordingly, one could look at the composition of the two groups and, in very concrete terms here, consider the stakeholder interests invested in these selection processes. For example, the Danish data ethics expert group had strong links to another governmental initiative, The Disruption Council, with several of the experts also active in this initiative. In addition, 8 of the 12 members of the group came from the private sector. Moreover, although the initiative was presented as an expert group on data ethics established by the Danish government, the work of the group was framed by the Danish Ministry of Industry, Business and Financial Affairs as an initiative to support Danish companies in turning data ethics into a competitive advantage.5 The HLEG on AI and its deliverables, on the other hand, received criticism for a composition that prioritised the ethicists and employers of AI effecting an “engineering-centred approach to ethical

issues” (Veale, 2019, p. 4) as well as industry stakeholders (Vasse’i, 2019). In particular, industry interests were described in an op ed in a German newspaper by one group member in terms of the “group’s extreme industrial weight”, which according to this member had directly affected “a lukewarm, short-sighted and deliberately vague” set of ethics guidelines (Metzinger, 2019).

We might also consider the balance of resources available to different members and stakeholder groups participating in the work of the group. For example, while some members of the HLEG on AI would have institutional resources available from the organisations or industries they represented, in terms of research and secretarial support, others – most often from civil society, including myself – did not have this type of support to help us in the highly demanding and time-consuming work involved in developing the deliverables of the Group. As such, a structural power imbalance was effectuated in a space in which each comment, statement and participation matters.

It is critical that we are vigilant of the different interests that shape these processes, especially in a time where stories about big tech industries’ lobbying of policy and investments in academic research were emerging on an almost daily basis. However, what I also realised during this process was that power is not a uniform interest expressed in coherent group formations and stakeholder groups, such as ‘the industry’, ‘civil society’ or ‘state’ stakeholder groups. Rather – as I therefore also explore in more detail in this study, and in particular in the second article on culture and AI – it is expressed in a complex set of systems of meaning making, ‘styles’, ‘world views’ and cultures. Exactly due to my immersion in the work of the HLEG on AI, I could compare the public criticism of stakeholder power dynamics with my own experience of the very negotiation in the group of concepts and statements of the different deliverables; in particular, I could consider this in the context of the many years I had participated in similar multistakeholder initiatives. Subsequently, I became observant of not only industry interests but also of interests in general that did not serve what I, from my own position, defined as the ‘human interest’. Company interests, organisational interests, individual EU member states’ interests, and at times even individual people’s personal interests in formulating specific ideas and conceptions of AI, human society and the role of human beings and technology. Thus, the power dynamics of this very process were much more
complex than the clear-cut interests between civil society and industry. Even within the groups of civil society representatives there were power struggles, and of course also within the industry. In this way, my immersion in the field guided me towards an investigation of more general cultural patterns of power negotiations and positioning that very often formed alliances across the more traditional interest formations and stakeholder groups (I describe these in the third article of this study on data interests and AI). Therefore, in the work of the group I also actively advocated a European cultural positioning of the ethics guidelines with an emphasis on the human-centric approach rather than a synthesis of all the ethics principles created at the time worldwide advocated by others. In addition, based on a discussion with two colleagues at the University of Copenhagen, Sille Obelitz Søe and Karen Louise Grova Søilen, I proposed the use of the term “Trustworthy AI” rather than the initial title of our work “Trusted AI” to emphasise the discussions in the HLEG on the accountability and implementation of AI ethics components. This is only an example of the essentiality of the immersive and reflexive approach of this study. As a researcher I was not “… neutral, detached, apolitical…” (Bordieu & Wacquant, 1992, p. 51) and I did not intend to be. Time and again, my research would be “bending back” on itself based on a reflection on the history, conditions and power dynamics of the research I conducted (Bordieu & Wacquant, 1992).

In addition to these initiatives of a more continuous character and with reappearing participations, I travelled all over Europe in the course of the study to speak and participate in debates at industry, governmental, intergovernmental and civil society conferences and events. These included the more traditional internet governance events such as the UN Internet Governance Forum (IGF) and the Computers, Privacy and Data Protection (CPDP) conferences, but also invitation only meetings and events, such as with the Club of Three, the OECD, the German Marshall Fund, closed and open meetings and activities of the Mozilla Foundation during the time of the development of their Trustworthy AI agenda, as well as many other industry, engineering and policy initiatives. At each of these events and initiatives, I experienced the unique power dynamics specific to the particular settings of the meetings and events, which included the

---

6 Sille Obelitz Søe and I delineate our thoughts on Trustworthy AI in this blog posting on DataEthics.eu Why Trust in AI is not enough (3rd September 2019). [https://dataethics.eu/why-trust-in-ai-is-not-enough/](https://dataethics.eu/why-trust-in-ai-is-not-enough/)
formulation of problems and solutions in the context of the formation of new and old power dynamics and interests. For example, at a Club of Three general assembly in London, the fortification and sense-making of historical cultural power alliance between France, Germany and the UK and their role in shaping the global agenda on AI were sought in the context of an upcoming UK Brexit. Another example was at the German Marshall Fund’s Brussels Forum where a perceived Chinese challenge was shaping discussions in particular. However, I gradually realised that all of these different spaces of negotiation were also part of the negotiation and formation of more global macro patterns of cultural positioning. I saw patterns of emerging alliances and formations across traditional interest groups. This was also where my investigative focus on data interests and data cultures emerged as organising principles and themes for my analyses, which would cut across interest groups and expressed in what I saw as increasingly shared reference systems of meaning and practice. For example, here I also unearthed the shared human-centric/centered approach (or what I in this study refer to just as the “human approach”) and its cultural references and history, when delineated in a European fundamental rights framework and even a somewhat vitalist perspective on humans and our tools and crafts, as particularly relevant for the formative theory of a ‘data ethics of power’ that I formulate in this study. For example, I attempt to explicate this in practical terms in the third article with a data interest analysis as a key reference framework in which various data interests can be negotiated and conflicts resolved in the interest of the human being.

Although I do not refer to these interactions (many confidential) in the study, they have also been formative for its general conclusions, arguments and call for action. They all emphasised my rising understanding of emerging cultural positioning and power dimensions and were directly responsible for the formation of my developed framework for a ‘data ethics of power’. In essence, I wanted to take power out of the papers and documents, statements and actions, as well as our practices in governing and acting on and with data, and make it into the ‘material’ of data ethics. As such, the next step would indeed be to apply this formative framework on data, power and interests with case studies of data systems and infrastructures, policy initiatives and documents as well as the communities and group formations and alliances that shape the development and governance of data and AI.
1.5 Interactions with Other Participants in the Research Field

From the outset, I generally considered my embeddedness in the field of study an important research value in that it helped me gain access to people and processes I might otherwise not have. However, my immersion in the field also presented the core research ethics challenges of this study. I therefore had to design a transparent research process with reflections on the choices I made. First, I addressed questions regarding the embedded bias of my research and power relations in my interactions with other participants in the field (Williamson et al., 1977). In particular, I took into account two aspects: Firstly, I had to consider in detail my position in the field and the research approach this would entail. I was aware that I did not set out to explore a phenomenon on its own terms, I had a preconception of the field that I wanted to explore and I had an action-oriented agenda. For this reason, I did not want to conduct a traditional ethnographic study of a group or community, nor did I consider it to be a case study as such. I therefore formulated the approach as a type of applied ethics with a specific objective and ontological point of departure (delineated in the last chapter of this study).

Secondly, I had to carefully consider how I engaged with other participants in the field of study as a researcher. As such, from the outset I made a reflective distinction between myself as an academic researcher and myself as the active participant in the research field. Accordingly, while my position in the field was indeed a strength in terms of access and in-depth understanding of particularly the historical dynamics of the field, it at times also prevented me from doing what other researchers might have done. For example, only in the first article do I include observations or notes from experiences, but I realised that this was compromising my reflective distinction between myself as a researcher and myself as a participant in the field. Accordingly, I decided not to include notes or observations from meetings or conversations with people in the rest of the study, simply because in those moments I was not present as a researcher. I was one of the people whose actions I would later reflect upon as a researcher.

As an active embedded participant in the field of this study during 2017–2019, I had many informal conversations with engineers, policymakers and experts. Although these have been key to the formation of my
understanding of the field, I do not reference them directly as I did not present myself as a researcher in these contexts. However, for the purpose of this study, I selected 14 people based on their expertise, and/or seniority in the field and historical experience whom I questioned about different aspects of my research. Some of these interactions are referred to directly in the first article of this study. The participants were specifically informed about the study and asked to consent to the temporary storage of personal data and the use of our interactions in anonymous form. These interactions do not form the empirical basis of the study, but are used in combination with the analysis of reports, participations and observations, documentations and public statements to clarify research questions and further nuance my depiction and understanding of the field.

The 14 people I interacted with as a researcher are listed as follows:

- Two policy advisors
- Four European institution officers
- One data protection commissioner
- One representative of a European country to the Committee of Ministers of the Council of Europe
- One European parliamentarian
- Two national ministry officers
- Three AI developers

The 11 decision-makers, policy advisors and civil servants contributed to my understanding of the policy-making dynamics by sharing their experiences with data/digital ethics in European policy-making. The conversations I had with them were between 20 and 40 minutes. I would primarily have them in breaks at events and therefore the time for interaction was limited. Conversations were prepared with semistructured guides, where I used a set of prepared questions and themes as point of departure, but did not limit the conversations to these questions. The preliminary guiding questions were as follows:
When did you start thinking about data ethics? Why? What does data ethics do? What is the role of data? Are policymakers aware of the role of data? Which policy areas are addressing the role of data? What is the role of legislation? What is the role of industry/developers? Where do we need legislation? Is data ethics different from data protection legislation? How? When, how and why did data ethics become part of the policy debate? What is the role of law/the GDPR? Who is responsible? What is the role of the EU?

I interacted with the three AI developers at the end of this study to clarify my understanding of the data interests involved in the design phase of AI.

At the beginning of the study, I combined the interactions with people with viewings, readings and analysis of reports, panel debates and presentations from the period 2015–2017. In particular, I focused on their narratives about data ethics, humans and technology and their cultural representations in terms of metaphorical structures of their arguments. Lakoff and Johnson (1980) famously described how our conceptualisation of the world is structured metaphorically. Understanding one thing in terms of another is not just an arbitrary exercise, they argue; metaphors give shape to and guide our actions, and they highlight or hide and crucially direct our focus (Lakoff & Johnsson, 1980, p. 10). Inevitably, metaphors “...play a central role in the construction of social and political reality” (Lakoff & Johnsson, 1980, p.159). However, more importantly, we make sense of the world of things “...on the basis of our own motivations, goals, actions, and characteristics” (Lakoff & Johnsson, 1980, p. 34). Hence, language is powerful; it naturalises our views of the world that enforce power dynamics, as Roland Barthes (1972/1957) illustrated in his analyses of the myths of human language in all its communicative forms. We may conclude then that the language we use to describe AI innovation, law, culture and governance also has a ‘governing’ function. That is, it shapes what we think we can do with the technology and how we can design, govern and direct it. Thus, I organised my analysis in a scheme with the following questions and themes in which I added the quotes and arguments from different participants in debates and texts from reports selected that would represent each of these metaphorically: What is data ethics? What is the human? What is technology? What is the challenge? What is data? How can we practically ensure this? What is the role of industry? What is the role of law and policymakers? What is the role of the EU?
Data ethics spaces of negotiation are possible when controversy arises.
In this first part of the study, I examine power and big data. I describe sociotechnical infrastructures based on big data technologies, which I refer to as big data sociotechnical infrastructures (BDSTIs) or “big data artificial intelligence sociotechnical infrastructures” (AISTIs; an evolution of the analytical capabilities of BDSTIs, which I describe in more detail in the second part of the study). In the 2010s they were increasingly representing and constituting global societies and environments as the mundane background against which social practice, social networking, identity construction, economy, culture and politics were conducted. They were in part institutionalised, in systems requirements standards for IT practices and in regulatory frameworks for data protection, they were invested with politics and human imagination about the challenges and opportunities of big data, and they were up for negotiation and contestation.

I explore the BDSTIs of the Big Data Society as a particular type of architecture of power for human agency and experience. They are constituted as material global networks that enable data collection and access across geographic territories and jurisdictions, and as flows around which dominant societal functions are increasingly organised (Castells, 2010). Accordingly, to design and shape the infrastructural components of BDSTIs is also an essential form of power.

A ‘data ethics of power’ addresses the new conditions of power of these technologically mediated configurations of space and time and their ethical implications. For example, BDSTIs’ spatial organisation of power is created for and by new types of dominant “managerial elites” (Castells, 2010, p. 445). Traditional arbitrary surveillance powers of states are therefore augmented by the powers of commercial actors that design BDSTIs to accumulate, track and access big data (Hayes, 2012; Pasquale, 2015; Powles, 2015–2018; Zuboff, 5 March 2016, 9 September 2014, 2019). As a result, surveillance powers of states and commercial actors alike are embedded in BDSTIs as a key property of their architecture and design (Haggerty & Ericson, 2000; Lyon, 2001, 2010, 2014, 2018; Hayes, 2012; Bauman & Lyon, 2013; Galic et al., 2017; Clarke, 2018).

I present sociotechnical infrastructures as social spaces occupied by interests in constant negotiation (Lefebvre, 1974/1992; Harvey, 1990) and
accordingly infrastructural practices as expressions of different political and cultural interests. In Europe, the idea of a ‘European infrastructure’ has generally been invested with the imagination and interest of an EU project that was to enable the efficient workings of a union of collaborating member states. That is, infrastructural practices, such as engineering and design standards, construction, investment and regulation are defined as a strategic endeavour to produce a space that enables a European economic and social union and community. This political aspiration has also been invested in the idea of a European digital single market translated in the 2010s into an aspiration to create European BDSTIs. The European BDSTIs are first and foremost defined here as differentiators on a global competitive digital market. Aspirations to compete in a global big data economy while preserving and protecting Europeans' fundamental rights were in the late 2010s reconciled in what has also been referred to as the European “third way” (with a particular emphasis on the development of European AISTIs). As such, the power of the big tech commercial elites of the global BDSTIs were increasingly challenged in Europe with concrete infrastructural practices, such as investment and policies supporting the development of European practitioner and user competences, science and research, technical data structures and data pooling as well as the development and possible implementation of legal frameworks to ensure the development of European BDSTIs and AISTIs.

Through exploring the various infrastructural practices of the 2010s invested in the development of BDSTIs, we see a moment of conflict and negotiation between the conception of a European cultural space and BDSTIs, AISTIs and global BDSTIs, and AISTIs. In this first part of the study, I argue that a ‘data ethics of power’ has a crucial role to play in critical moments such as this, as they lead to the cultural compromises or “technological momentum” (Hughes, 1983, 1987) that a global sociotechnical system needs to evolve. They are also crucial to phases of innovation and development as they constitute the very transformation of the sociotechnical system that emerges out of a quest to solve critical problems of the system. Data ethics can therefore also be identified at a meso level of analysis in institutions, companies, governments and intergovernmental organisations that either move along the natural state of affairs or establish initiatives and practices, namely ‘spaces of negotiation’, dedicated to the value and ethical reflection that must accompany legal compliance.
CHAPTER 2
Big Data Sociotechnical Infrastructures (BDSTIs)
FROM 1967 TO 2018, the Morandi Bridge ran as one of the main arteries through Genova, Italy, connecting the East to the West part of the city. The bridge’s concrete construction was a global symbol of Italian engineering and technical capacity. In fact, Italy was in 2018 one of the top cement producers in the world and thousands of concrete viaducts, tunnels and bridges worldwide were based on the Italian design. As a key infrastructural component of the city, Morandi Bridge formed the silent background against which life and business were facilitated with thousands of people driving across it every day without giving the bridge and its construction an extra thought. However, on the 14th of August 2018, the heavily trafficked bridge collapsed, causing the death of 43 people and leaving 600 homeless. The bridge was at once no longer silent. With countless media reports and investigations, not only was the engineering history of the bridge told but also the very infrastructural breakdown was equated with the collapse of a “national myth” (Mattioli, 2019).

Think of BDSTIs like the Morandi Bridge, or just any ordinary road and building that reside in a “naturalised background” (Edwards, 2002, p. 185). We cross them, like we cross bridges and follow roads, every day. Unnoticed, they facilitate and organise our everyday lives. They constitute the micro spatial architecture of our everyday lives and they are embedded in macro societal structures. And just like the Morandi Bridge, these BDSTIs are not just appearance, digits and cables; they have politics and culture that become particularly apparent in moments of crisis and infrastructural breakdown (Star & Bowker, 2006; Bowker et al. 2010; Harvey et al., 2017).

In this chapter, I examine the infrastructures of the Big Data Society with the double purpose picked up from studies of societal infrastructures to know and make transparent a human environment, but crucially also to control it (Harvey et al., 2017, p. 2). The main objective is to understand the special power dynamics of the sociotechnical infrastructures that a data ethics of power addresses. I refer to these as BDSTIs or AISTIs and investigate how they have evolved as components of the power dynamics of the Information Society addressing specifically the conception of a European big data infrastructure. In the last part of the chapter, I consider
the ethical problems that concern a data ethics of power specific to the Big Data Society, which is the term I have chosen to use to describe the specific characteristics of societies in which BDSTIs and AISTIs are dominant.
BIG DATA SOCIETY

What is a Big Data Society? Society’s technologically advanced big data technologies and systems aside, how do we imagine the role of big data in a society like this? What social, economic and cultural functions should it have? If we consider the Big Data Society a coherent social structure, we also need to understand its social, economic and not least ideological underpinnings. Mayer-Schonberger and Cukier depict the Big Data Society as a societal revolution that transforms human work, social relations and the economy. This is a transformation enabled by computer technologies and dictated by a transformation of all things (and people) into data formats (“datafication”, Mayer-Schonberger & Cukier, 2013, p.15) in order to “quantify the world”, thus helping businesses, governments and scientists organise and make sense of data (Ibid., p. 79).

This evolution of BDSTIs can be coupled with the imagination of big data as an unlimited resource that, unlike other resources in society (e.g., oil and water), will not diminish (Ibid., p. 101). In essence, we may therefore also argue that big data more than anything is also a movement behind which lies a system of imagining and making sense of the role of digitalised data in society (Mai, 2019, p. 111). The collection of big data is here perceived as an end in itself, holding the promise of future endless ways of use and reuse (Mayer-Schonberger & Cukier, 2013, p. 100). The limits and risks are only technical in nature: limited storage, processing and analytical capacities. Accordingly, the most powerful companies and institutions are also those with a “big data mindset”, engaging in big data infrastructural practices, collecting big data, and processing and creating interoperable big data sets (Ibid., p. 129).

These ideas about the risks and potentials of big data can be traced back to the late 1990s, when big data surfaced as a term in the computer science and business fields to describe a range of emerging technological innovations in digital data storage, exchange and analytics enabled by computer technologies and the evolution of the internet. It was first used by a chief scientist named John Mashey at Silicon Graphics, a large U.S. computer graphics company, in a number of product pitch talks depicting the great promise of big data, but also describing the commercial and technical challenge to meet this future potential (Lohr, 2013). For example,
in 1999 he predicted how big data would unsettle both human and material IT infrastructures. The response was an urgent call for companies to “Change: survive!” as one of his PowerPoint presentations exclaimed in 1999, with reference to among others enhanced computer power to store and process data and the unleashing of data with scalable interconnect and high-performance networking (Mashey, 1999, p. 45).

2.1 The Imagination and Politics of Space and Infrastructure

The BDSTIs of the 2010s were human-made space shaped by commercial and institutional fantasies about the potential of big data as an unlimited resource and the commercial and technical risks to companies and infrastructures that failed to store, collect and process it. It was therefore also shaped by practices aimed at making the most of this potential while simultaneously mitigating the perceived risks.

The Marxist philosopher and sociologist Henri Lefebvre depicted space as a composite of a material physical reality and social practice, or a type of space that does not exist without “…the energy that is deployed within it” (Lefebvre, 1974/1992, p. 13). He divided this social energy invested in space into three types: “the perceived, the conceived, and the lived” (Lefebvre, 1974/1992, p. 39). In other words, we perceive space physically with our perception, and we feel space qua our positioned bodies; however, space is also conceived by, for example, urban planners, engineers and scientists and it is lived with imagination that seeks to “change and appropriate” it (Lefebvre, 1974/1992, p. 31). Thus, he pointed to struggles over the meaning of space, delineating the power dynamics and politics that shape space as a real and imagined resource invested with specific interests. Generally, space is open for active “occupation” by interests. For example, “global space”, he argued, was originally established in abstract as “…a void waiting to be filled, as a medium waiting to be colonized” (Lefebvre, 1974/1992, p. 125).

Space is not just material in nature. It consists of open plains and our imagination invested in these voids of space, and also its materialisation in human-made infrastructures. These infrastructures are in essence the navigational tools and architectonics of our everyday lives. They physically direct us in specific directions and limit us corporeally. One cannot walk through a wall or cross a border without showing a passport. Similarly, the social or cultural dimensions of space can limit or create opportunities. I
cannot cross any border with just any passport, and even if I manage to cross the first human border controller with my passport, I might be stopped at the next electronic border when my face is recognized and correlated with a face in the crowd of people in a public demonstration.

In this way, the very design of an infrastructure is an active process. We do not just fill in a void of open space with an infrastructure—‘we infrastructure’ as the information and STS scholars Geoffrey C. Bowker and Susan Leigh Star calls it (Star, 1999, Star & Bowker, 2006, Bowker & Star, 2000). That is, practices of designing, repairing or even participating and navigating within infrastructures are also active participation in social power dynamics. As such, an infrastructure is a design, an engineering process and a mere presence, but it also has politics, imagination, culture and power dynamics (Larkin, 2013).

We can equally think of BDSTIs as *spaces*. They certainly are perceived and lived as such. We are positioned in their architectures, and life opportunities are found or lost through this architectural positioning. More critical is that these BDSTI spaces are not a natural given. They are occupied space, imagined in terms of the risks and opportunities of a “big data mindset” and shaped accordingly. This is a liberating recognition, because in this way we can also think of the space and architectonics of BDSTIs as something that can be reimagined and rebuilt.

Lefebvre introduced two types of space, namely the space of social practice, or what he referred to as “real space”, and the “ideal space” (Lefebvre, 1974/1992, p.15). They are not separate, he argued, but intertwined, each presupposing the other. In other words, space is a composite space of its intended purposes and the social negotiation of these. Similarly, we may consider the infrastructures of the space we navigate as dynamic “…extended material assemblages that generate effects and structure social relations, either through engineered (i.e. planned and purposefully crafted) or non-engineered (i.e. unplanned and emergent) activities” (Harvey et al., 2017, p. 5).

However, in saying this, we can also argue that infrastructures, like BDSTIs, are always sites of power struggles over different imaginations and hopes regarding the social appropriation of space (Larkin, 2013). In history, the development of spatial infrastructures often produces social conflicts (Reeves, 2017). Winner (1980) for example famously used the low hanging overpasses that connect Long Island to the rest of New York as a case in
point. They were specifically designed to prevent access to public buses that were used by the majority of black people, and thus, Long Island could mainly be accessed by middle or upper-class white people with private cars. In this way, one social group was prevented by infrastructural design from accessing the recreational areas of Long Island.

Infrastructures are “narrative structures” of social power. A study of infrastructure should therefore, as Susan Leigh Star argues, always seek to restore these types of social narratives (Star, 1999, p. 377). Making the invisible factors of infrastructures visible by pulling the underlying (the ‘infra’) into the foreground also has a social function (Star, 1999; Bowker & Star, 2000; Bowker, 2010). It makes change possible and the hidden social consequences manageable (Bowker et al., 2010, p. 98). Nevertheless, infrastructures are most often unquestioned as long as they run smoothly. However, when disrupted like the Morandi Bridge, a more detailed explication of their inner working is required (Star, 1999). We may use here a disruption of the 2010s BDSTIs to explore their infrastructural narratives.

In the early 2010s, BDSTIs were intrinsically intertwined with organisational systems and practices (Ratner & Gad, 2019). They were sustaining a market and commercial structure and had evolved into sociotechnical infrastructures for the flows of global economies and societies, cutting across legal jurisdiction, cultures and societies. BDSTIs and the imagination and corresponding practices invested in these were representing and constituting global societies and environments as the mundane background against which social practice, social networking, identity construction, economy, culture and politics were conducted. Yet in 2013, the Snowden revelations of the U.S. National Security Agency (NSA)’s global mass surveillance system pulled BDSTIs to the foreground of public debate, and they were exposed in all their complexity. The PowerPoint presentations from U.S. intelligence officers detailing the PRISM program that were leaked and published in the Guardian (theGuardian, 2013) revealed among other things the mass surveillance intelligence system’s intertwinement with the largest global big data companies’ social networking services. They also provided a detailed map of the world’s (Europe, U.S. and Canada, Latin America and the Caribbean, Asia & Pacific, and Africa) flows of data and communication directed through nodes and hubs in the US. Phone calls, emails and chat would not take the most physically direct path, the PRISM slides stated, but rather the cheapest,
which would go through the US. Collection of data could first be conducted through fibre cables and infrastructure as data flowed past, but also directly on the servers from US-based companies servicing world users with, for example, social networking services. In this way a simultaneously material and immaterial transinternational infrastructure had enabled the mass collection of data of European citizens by a foreign intelligence service.

These revelations among others resulted in the EU scrapping the Safe Harbour agreement between the EU and U.S., which had previously provided the legal framework for the data exchange infrastructure between the US and the EU market. However, what became specifically evident with Snowden’s disruption of the global big data infrastructure was its highly concrete and detailed illustration of a new big data global infrastructure’s profound challenges to traditional modes of territorial governance of world states and regions. In the years following these revelations, news of a range of other cases of infrastructural ‘disruptions’ of the big data infrastructure was revealed from large data hacks and leaks of companies and institutions (for example, the Snapchat hack in 2013 or the Ashley Madison hack in 2015; list of breaches on Wikipedia, 2020) to the revelation of the complex data analytics used to influence electoral votes (Cadwalladr, 2017; Rosenbergh et al., 2018, theGuardian, 2018).

These incidents all caused momentary or long-lasting disruptions to existing governmental and business BDSTI practices that had been imagined, conceived of and implemented since the early 2000s. In particular, as I argue in the first article of this study on data ethics in European policymaking, due to this disruption and crisis of the BDSTIs, the big data imaginations and mindsets of the early digital developments were in the late 2010s increasingly contested by alternative mindsets and imaginations about the conception and implementation of a European digital infrastructure. In the following section, I illustrate how the concept of a European BDSTI emerged and evolved in the 2010s in the context of different competing imaginations and narratives about the potential and risks of big data.
2.2 The Narrative of a European Digital Infrastructure

In European policy-making, the term ‘infrastructure’ is first and foremost used to describe a system across Europe that enables cohesion and collaboration, including the movement of people and goods, between member states. In that way, we might consider this ‘European infrastructure’ a space occupied by the imagination and interest of an EU project that enables the efficient workings of a union of collaborating member states.

The EU was historically created after the Second World War as an economic collaboration between countries, and the creation of a single market was sought. It later evolved into a political union around areas such as foreign policy, migration and security. According to this political project, a European infrastructure’s architectonics should facilitate first and foremost a European union of member states; that is, European citizens’ free movement as well as the European single market’s free movement of goods and services. For example, the Trans-European Transport Network (TEN-T) policy was created with the objective “...to strengthen social, economic and territorial cohesion in the EU” (European Commission A, 2020) and concerns the implementation and development of a network of the EU’s physical transport infrastructure, which in 2017 counted over 217,000 km of railways, 77,000 km of motorways, 42,000 km of inland waterways, 329 key seaports and 325 airports (European Commission B, 2020). The TEN-T network is part of a system of Trans-European Networks (TENs) that also covers energy and digital services. These programs are supported and implemented via the Connecting Europe Facility, which is a 30 billion euro (in 2019) funding instrument in the form of grants, procurement and financial instruments with the stated aim to “further integration of the European Single Market” (European Commission C, 2019, p. 6). The European infrastructure is here defined in terms of its political purpose to support an imagined European community; that is, to do infrastructure in the EU is also perceived as a strategic endeavour, and from this initial political intention emerged infrastructural practices accordingly, such as engineering and design standards, construction, investment and regulation, which produce space, so to speak—that is to say, the engineered and intended components of a European infrastructure.

In the 2010s, an effort to extend the material infrastructure of the EU with a digital sociotechnical infrastructure was increasingly voiced in EU official
strategies and materialised in infrastructural practices, such as dedicated policies and investments. The EU’s Digital Agenda intersected policy areas and regulatory frameworks that traditionally were treated separately (Valtysson, 2017). The term infrastructure was here still only used to describe the technical aspects of a digital infrastructure for the single market. However, although not described as such, the social and economic components of the European digital infrastructural architectonics gradually became a focal point of European policy and investment strategies.

In 2010, a Digital Agenda for Europe was presented with the description of a “Digital Single Market”: “It is time for a new single market to deliver the benefits of the digital era” (European Commission D, 2010, p. 7). Within the agenda, the Digital Single Market aspiration was voiced in a concern over a persistent fragmentation that was said to restrain Europe’s competitiveness in a digital economy overshadowed by companies such as Google, eBay, Amazon, and Facebook that “originate outside the EU” (European Commission D, 2010, p. 7). A wide spectrum of infrastructural practices were therefore envisioned to develop this competitive space of the Digital Single Market. Some were technical in nature, focused on “interoperability” and the creation of technical standards, or the development of “fast and ultra-fast internet access”. However, immaterial components were also described, such as ensuring the “trust” of Europeans: “the digital age is neither ‘big brother’ nor ‘cyber wild west’” (European Commission D, 2010, p. 16).

Later in 2015, a Digital Single Market Strategy for Europe to further enable the single market “…free movement of goods, persons, services and capital…” was published (European Commission E, 2015, p. 3). In his introduction to the strategy, the then-president of the European Commission Jean-Claude Juncker spelled out its foundational political purpose and imagining: “I believe that we must make much better use of the great opportunities offered by digital technologies, which know no borders. To do so, we will need to have the courage to break down national silos in telecoms regulation, in copyright and data protection legislation, in the management of radio waves and in the application of competition law” (European Commission E, 2015, p. 2).

This was followed up with two communications published together in 2016 (“Digitising European Industry Reaping the full benefits of a Digital Single Market”, European Commission F, 2016 and the “European Cloud Initiative - Building a competitive data and knowledge economy in Europe”,
that each emphasized the impact and role of big data and spelled out the contours of a European BDSTI.

In the first communication, big data was described as the foundation of an industrial revolution, and a new focus on a data sharing cloud-based infrastructure for scientists and engineers in the EU also took form (European Commission, F, 2016 p. 2). It outlined concrete infrastructural practices to develop a data infrastructure in terms of investment, policy and coordination.

In the second of the two 2016 communications, a concrete infrastructural initiative to support the development of a “European Cloud” with an emphasis on a “European Data Infrastructure” (European Commission, F, 2016, p. 8-10) was presented, and the BDSTI was specifically framed as a promise to strengthen a data and knowledge economy in Europe and use the potential of big data. It was first and foremost presented as a data-sharing system consisting of different technical components: “the data infrastructures which store and manage data; the high-bandwidth networks which transport data; and the ever more powerful computers which can be used to process the data” (European Commission, F, 2016, p. 2). However, the European Cloud was not just a technical data infrastructure and organisation; it was also envisioned to facilitate the potential of big data by making “it possible to move, share and re-use data seamlessly across global markets and borders, and among institutions and research disciplines” (European Commission, F, 2016, p. 2).

Following these first depictions of a European data infrastructure, however, the aspiration to become a European digital single market and data infrastructure competitor in a global market on similar terms as big data companies gradually transformed into an aspiration to make this European data infrastructure and digital single market a key differentiator in a global competitive digital market. The aspirations to compete in a global big data economy while preserving and protecting Europeans’ fundamental rights were soon reconciled in what was also referred to as the European “third way”, which I describe in more detail in the second article of this study on culture and AI. Like so, in 2020 a general “European data strategy” was proposed with reference to a society “empowered by data” and recognizing the role of data in society: “Data will reshape the way we produce, consume and live. Benefits will be felt in every single aspect of our lives, ranging from more conscious energy consumption and product, material and food
traceability, to healthier lives and better health-care” (European Commission, H, 2020, p.2)

As such, concrete infrastructural practices were outlined, such as investment and policies supporting the development of practitioner and user competences, European science and research, technical data structures and data pooling as well as the development and possible implementation of legal frameworks, to ensure the development of European digital data infrastructures.

2.3 The Power Infrastructure of the Information Society

In the early 2000s, the ‘Information Society’ was an established term and forceful strategic policy focus, not only in Europe but also on a global scale. Most significantly, this political agenda gained footing during the UN World Summit on the Information Society (WSIS). In 2003, the first summit was held in Tunis with the stated aim of gathering momentum and taking concrete political steps to establish the foundations for an inclusive Information Society and “...reflecting all the different interests at stake” (World Summit on the Information Society, 2013). Based on a sense of urgency and realisation that an ongoing digital revolution was transforming society as we know it, governments from around the world gathered at one of the UN’s first multistakeholder fora to create a political agenda with the aim of tackling the societal, economic and cultural implications of a rapidly developing sociotechnical Information Society.

Frank Webster (2014) describes preoccupations with the Information Society as a prioritisation of information. Even though there are different views on what this prioritisation of information means, and what role information plays in societies, it is fundamentally, he argues, a way of conceiving something new and different about contemporary societies (Webster, 2014, p. 8). He examines the literature that describes the information society and finds five definitions, not necessarily mutually exclusive, but each emphasising different aspects of the new role that information is envisaged to play in society at the end of the 20th Century and beginning of the 21st. Technological definitions are concerned with the evolution of specific ‘new technologies’ in society, such as the computer and IT technologies in general. Economic definitions focus on the economic value of informational activities, whereas occupational definitions track an
increase in informational work, spatial definitions examine the role of information networks in reshaping space and time, and cultural definitions of the Information Society look at the increasingly media-laden society and technological information environment (Webster, 2014, p. 10-23).

Here, I want to introduce a different perspective on the particularities of the Information Society. Rather than a prioritisation of information, we may consider all of these different definitions of what constitutes the Information Society horizontally in terms of a redistribution of power facilitated within new technologically mediated configurations of space and time, or what I refer to as BDSTIs and AISTIs. In the following, I will illustrate how.

As a point of departure, I want to explore the technological evolution of the space for not just the transportation of information but also the architecture of global systems of power. Stephen Kern (1983) illustrated how technological developments between 1880 and 1918 expanded our experience of space. With the telephone, one could be in two different places simultaneously. With the wireless, this simultaneity of experience was expanded to an instant sense of the whole world, and only a few decades after Lefebvre’s 1974 (1992) description of an occupied global space of commercial images, signs and objects, the digital evolution of a global space was complete. The Google Maps service has, for example, transformed geographical space into a digital data infrastructure with satellite imagery, aerial photography, street maps, and 360° panoramic views of streets combined with real-time traffic conditions and route planning.

Looking at the 21st Century’s technological evolution, which Google maps is representative of, sociology scholar Francesco Lapenta (2011) coined the term “Geomedia” to describe emerging location-based services like Google Maps and Earth’s merging of geographical space, virtual space and the local experiences of users based on big data and information exchanges. He describes these as mediating spaces that function as “new organisational and regulatory systems” articulating and organising social interactions (Lapenta, 2011, p.21). They are used by individuals as social navigation tools that can help reduce the complexity of global information systems to manageable and socially relevant information exchanges (Lapenta, 2011, p. 21). Geomedia is an example of technologies that in the 21st Century Information Society were transforming the body, social and individual experiences, physical space and location into interoperable digital data, blurring their lines of separation when integrated into the designed spatial architectures of a
virtual infrastructure. As such, not only our experience of space transformed, but Geomedia, as Lapenta argues, also regulate social behaviour and interpersonal communications as well as coordinate social interactions. In other words, they transform the very spatial architectonics, creating what we have previously referred to as a Big Data Society.

The cultural geographer David Harvey used the term “time-space compression” (Harvey, 1990) to describe the transformation of the human experience and thus representations of time and space in an increasingly globalised world, which had reduced spatial distances through temporal reductions. As such, he pointed to a concrete transformation of the very objective qualities of time and space presenting an image of a shrinking world map. He explained this change through what he called the annihilation of “space through time” (Harvey, 1990, p. 241); that is, a reduction of distances between places in terms of travel time and costs. This again he saw as the result of history and human thinking, originally the European 17th Century Enlightenment thinkers’ preoccupation with economic and political ‘production of space’: “The production of turnpikes, canals, systems of communication and administration, cleared lands, and the like put the question of the production of a space of transport and communications clearly on the agenda” (Harvey, 1990, p. 255).

Harvey argued that space is conquered by humans through these processes of ‘producing space’, of occupying and settling in space, and that these occupations of space are legitimised by specific legal systems that stipulate the different rights we have to the spaces we navigate in society. As, for example, as I illustrated in the previous section when the stipulation of European citizen rights was incorporated into the infrastructural components of an evolving European space in the 2010s. In this way, space constitutes the internal and external spatial borders of a society. These types of ‘occupied spaces’ frame social process and practices (Harvey, 1990, p. 258). Therefore, the transformation of time and space also has the function of maintaining power as it imposes the structure for social practices representing the forces of power in a given society. Harvey saw “time–space compression” and a shrinking world map as not just consequences of technological developments per se, but also importantly as an expression of the embedded interests of the expansion of capitalism and industrialisation in the 19th Century. Hence, while a 17th Century space was occupied with human ideas about a ‘better society’ and accordingly focused on a rational
ordering of space and time to develop a society that guaranteed individual liberties and human welfare, “time–space compression” is, he argued, mainly created for the operations of capital and therefore designed for instantaneity, ephemerality, fragmentation, volatility and disposability (Harvey, 1990, 286-307). In similar ways, I demonstrated before how the engineering goals for creating infrastructural conditions for the free flow of a single European market, in the 2010s existed side by side with aspirations to develop infrastructural components that protects European citizens’ fundamental rights.

I want to use Harvey’s description of space as open for “active occupation” by different ideas and interests to consider the infrastructural practices that actively contribute to the material and immaterial shape, form, direction and orientation of its power architectonics. This type of occupation works in a space of imagination and symbolic practices of power, and it works in very concrete terms as a property of the material space we design and create for ourselves in society. As follows, our ‘space’ in the early 21st Century was very much constituted by material forms of power imitating and structured in terms of the special characteristics of the global and local virtual spaces that were imagined in business and politics and designed accordingly.

Manuel Castells (2010) describes the IT revolution as a major historical event comparable to the 18th Century’s Industrial Revolution, and refers to the “space of flows” (of capital, information and technology, organisational interaction, images, sounds, and symbols) as the material form of power in a society where dominant societal functions are organised around networks (Castells, 2010, p. 407-459). Moreover, it is this very architecture of flows that constitute the transformation of power in society (Castells, 2010, p. 445).

There are three layers to this “space of flows”. The first layer Castells refers to as the “material support” constituted by a “circuit of electronic exchanges” in a global technological information network (Castells, 2010, p. 442), which also forms the foundation of an acceleration of the moment of people and goods. The second layer is constituted by “nodes and hubs” (Castells, 2010, p. 443); that is, the networks that enable the spaces of flows are not “placeless”, but rather organised around electronically linked up “places” with “well-defined social, cultural, physical functional characteristics.” (Castells, 2010, p. 443). They have specific functions such as exchange or communication “hubs”, or as the “nodes” where strategically significant functions are located in positions with different constantly
evolving hierarchies between them. The third material layer of the space of flows involves its spatial organisation and form created for dominant “managerial elites”. This takes its point of departure in the general notion that a society is “asymmetrically organised around the dominant interests specific to each social structure” (Castells, 2010, p. 445).

This particular three-layered “architecture” constitutes the foundation for an ongoing transformation of society, and it is in this very design and architecture that we may read “...the deeper tendencies of society, of those that could not be openly declared but yet were strong enough to be cast in stone, in concrete, in steel, in glass, and in the visual perception of the human beings who were to dwell, deal, or worship in such forms.” (Castells, 2010, p. 448). In Castell’s depiction, societal power is therefore concentrated in the very information architecture of technological networks. It is no longer fixed in places, but distributed in the design of infrastructures of information flows “... the power of flows takes precedence over the flows of power” (Castells, 2010, p. 500). Thus, to be connected or disconnected from the space of flows is the first step towards being empowered in the network society, while the second is to actively participate in the design and shaping of its global infrastructure.
ETHICS IN THE SURVEILLANCE SOCIETY

I have illustrated how power transforms in the networks and flows that comprise the architecture of the Information Society, and also how politics and narratives about big data are invested in the very construction of the infrastructures that facilitate the new type of sociotechnical power invested in a Big Data Society. I have not yet addressed the ethical problems, and thus have at this point no place to apply a 'data ethics of power'. To do this, I will in the following section therefore outline theories that portray the specific ethical problems of a society and power structure as such.

2.4 Liquid Surveillance

I consider the BDSTIs of the 2010s a form of power integrated in our spatial architectures. I believe that they are not liberating spaces in which human life flourishes. They sustain the power of powerful actors in society while putting others at a disadvantage, and they are difficult to resist and change particularly due to the very design of their data infrastructures, which track and monitor personal data by default and restrict citizens' liberty and agency. In other words, the acceleration and integration of – what I have referred to in another place as “Destiny machines” (Hasselbalch, 2015) – in ordinary state and business practices have resulted in a complex and advanced societal machinery that leads, guides and defines human lives.

Destiny machines are technological systems and processes designed to predict human behaviour based on the accumulation of personal data and then act on these predictions. Every day humans interact with these big data “machines” designed to predict human behaviour by tracking, scrutinizing and analyzing a recorded and stored data past and present of the “data doubles” of humans (Haggerty & Ericson, 2000). In this way, human lives are framed and pointed in specific directions. Destiny machines can also be described as machines specifically designed to produce machine-readable people and shell out destinies on the other side of the production line. In fact, they produce, create, act on and define destinies. We might even say that fate is what Destiny machines produce. This is what is being innovated with; it is part of an actual machinery and can literally be sold and traded.
with (Hasselbalch, 2015). Within this machinery, human lives are made meaningful only within the sorting structures of inclusion or exclusion of the surveillance assemblage (Lyon, 2010).

In fact, the Destiny machine does not need the human life and body; it is fundamentally indifferent to the individual human being as only our “data doubles” are meaningful within its surveillance assemblage. Or, said another way, it only has an interest in the “data derivatives” (Amoore, 2011) of the “data double”. As Louise Amoore describes, it is “not centred on who we are, nor even on what our data say about us, but on what can be imagined and inferred about who we might be – on our very proclivities and potentialities.” (Amoore, 2011, p. 24).

In the early 2000s, the sociologist David Lyon outlined the ethical challenges of what he referred to as the “Surveillance Society” (Lyon, 2001), or more specifically a “Liquid Surveillance Society” (Lyon, 2010, Bauman & Lyon, 2013), which is sustained by socio-technical “data flows” (Lyon, 2010, p. 325). Lyon’s accounts of the key properties of the power structures of the Information Society, and the field of Surveillance Studies that he initiated, should be considered the foundation for any ethical and generally critical concern with the role of big data sociotechnical systems and their economies, politics and cultures.

“Liquid surveillance” has a different shape than the form of surveillance described in Jeremy Bentham’s Panopticon prison (1787) and Michel Foucault’s Panopticism (Foucault, 1975/2018), which is centrally integrated in the spatial architecture of society and enforced as a type of aware self-discipline. It does not come from a centralised visible above (‘sur’), or middle as in the Panopticon prison (Baumann & Lyon, 2013), but is embedded in digital infrastructures, networked, distributed and sustained by increasingly greater distances between the ones that watch and those being watched (Galic et al., 2017). Opaque and bottom up, liquid surveillance is invisibly intertwined with individuals’ lives, and therefore, it is also inscrutable and difficult to address (Lyon, 2010). Notably, surveillance in the Big Data Society is not exceptional, but a condition of experience and human life. It is our culture (Lyon, 2018). Based on “dataveillance”, a systematic monitoring, tracking and analysis of personal data systems (Bauman & Lyon, 2013; Clarke, 2018). Surveillance takes form as an “assemblage” that abstracts the human body from a digital “data double”, which can be
scrutinized and used for purposes of control by governments or can be sold for profit in commercial interchanges (Haggerty & Ericson, 2000).

Understanding the transformation of the actors of power in the surveillance society necessitates a widening of ethical scrutiny from a more conventional attention to the arbitrary surveillance powers of states to the commercial stakeholders that gain power through accumulation, tracking and access to big data. Surveillance is a “surveillance-industrial complex” in which it is the very sociotechnical interrelation between state and private sector actors that makes surveillance possible (Hayes, 2012).

This change in power dynamics is a core ethical problem as it is based on an increasing information asymmetry between individuals/citizens and the powers of the big data companies that collect and process data in digital networks (Pasquale, 2015; Hasselbalch & Tranberg, 2016; Powles, 2015–2018; Zuboff, 5 March 2016, 9 September 2014, 2019). As Tranberg and I illustrated in our book (2016): “...the biggest risk lies in the unequal balance of power that the opaque data market creates between individuals and corporations.” (Hasselbalch & Tranberg, 2016, p. 161)

In particular, Shoshana Zuboff’s important work on “Surveillance Capitalism” drew public attention in the late 2010s to the concrete role of named powerful Silicon Valley industrial actors, such as Google and Facebook. Zuboff describes “Surveillance Capitalism” in her 2019 book of the same name as an accumulation of a capitalist logic that commodifies human psychology and experience to satisfy market forces and commercial aims of tech giants. Her main concern is the way in which these new commercial forms of digital surveillance reshape the institutional structure of modern democracies. She describes it in very concrete terms as follows:

“Two men at Google who do not enjoy the legitimacy of the vote, democratic oversight, or the demands of shareholder governance exercise control over the organization and presentation of the world’s information. One man at Facebook who does not enjoy the legitimacy of the vote, democratic oversight, or the demands of shareholder governance exercises control over an increasingly universal means of social connection along with the information concealed in its networks.” (Zuboff, 2019, p. 127).

It is in these studies of contemporary structures of power and surveillance that an urgent call for data ethical action concerning the power of big data
Lyon emphasises the urgency of developing an “ethics of surveillance” (Lyon, 2010, p. 333), or what he in conversation with Zygmunt Bauman also refers to as “surveillance ethics”, to address the “political realities of surveillance” (Bauman & Lyon, 2013, p. 20). Baumann and Lyon identify two major issues addressed by an ethics as such: one they refer to with Bauman’s term “adiaphorization” (Bauman, 1995), where morality is abstracted from the very systems and processes of surveillance, while the other is the distance created between the human being and the consequences of their actions (Bauman & Lyon, 2013, p. 7).

Lyon generally envisions a very practical applied data ethics, or what he refers to as an “ethics of Big Data practices” (Lyon, 2014, p. 10), to renegotiate what is increasingly exposed as being an unequal distribution of power between individuals and the institutions that develop the BDSTIs. Later, with direct reference to the 2013 Snowden revelations, he describes this as follows:

“We need ethical tools for assessing surveillance, a broadened sense of why privacy matters and ways of translating these into political goals. And it is essential that we do this with a clear sense of what kind of world we are working towards. How do we get a sense of what a better world would look like?” (Lyon, 2014)

### 2.5 Big Data Ethics

Studies of the surveillance architecture of the Big Data Society purposely seek to disclose new constellations of power in the sociotechnical assemblages of said society and, by doing so, to hold those in power accountable. Scholar of surveillance, Eric Stoddart, argues:

“As a discourse that discloses what is being done to us, surveillance studies itself is, although at first not obviously so, a method of ethical enquiry.” (Stoddart, 2012, p. 369).

To ethically evaluate surveillance, Stoddart considers two strands of approaches, the first of which is the “discursive-disclosive” approach, which he defines very clearly as an ethics that seeks “to disclose what is being done and the possibilities that might be available for alternative actions” (Stoddart, 2012, p. 372). He refers here to a type of Foucauldian ethics in which ethical inquiries address practices of surveillance, rather than only processes, and as such he emphasises ethics as a process of liberating reflection. As he describes it, “A discursive approach discloses to both us and others what we
did not previously know about our situations, the conditions under which we have been living and working and how we might be being exploited...” (Stoddart, 2012, p. 372).

The second approach Stoddart considers is what he refers to as a “rights-based” approach with reference to a body of human rights work with the aim of demanding “accountability of those with the power to watch” (Stoddart, 2012, p. 369). Here, it is specifically the challenges to human rights posed by the new power actors in, what has been referred to as the “platform economy”. It is an economy shaped by the large digital technology platforms of the 21st Century and that transform the structure for and accordingly adequacy of traditional modes of individual rights protection (Belli & Zingales, 2017; Wagner et al., 2019; Franklin, 2019; Jørgensen, 2019). Legal scholars Neil M. Richards and Jonathan King (2014), for example, argue the following when considering the right to privacy in the context of big data sociotechnical systems:

“Existing privacy protections focused on managing personally identifying information are not enough when secondary uses of big data sets can reverse engineer past, present, and even future breaches of privacy, confidentiality, and identity” (Richards & King, 2014, p. 393).

They thus suggest a more inclusive analysis based on a “big data ethics” (Richards & King, 2014, p. 393), which points to the ethical implications of the empowerment of institutions that possess big data capabilities at the expense of “individual identity” (Richards & King, 2014, p. 395).

Crucially, when addressing the distributed power relations of the Big Data Society as a condition for the implementation of the right to privacy in this way, we may also better understand privacy as “contextual” (Nissenbaum, 2010), effected and created in groups (Mittelstadt, 2017), and therefore a collective rather than only individual responsibility (Tisne, 2020).

From these depictions of the surveillance properties of a sociotechnical data-infused environment, I draw the very material of a ‘data ethics of power’. The spatial architecture of BDSTIs, and also AISTIs, is by design sustaining asymmetries of power, but they are not just reinforcing existing power dynamics—they are also creating new structures and actors of power. Commercial actors gain power with the accumulation of data and data design made for purposes of profit and/or control of challenging traditional state power.

A ‘data ethics of power’ addresses this new natural state of power in the Big
Data Society and calls for an alternative design and implementation of data systems; however, it also importantly calls for different data cultures (a term I return to in the second part of this study). Here, “surveillance capitalism” (Zuboff, 2019) is an interesting term that incapsulates the role and power structures of capital and commercial actors very well. Yet, it does not capture the “liquidity” (Baumann, 2000; Bauman & Haugard, 2008; Lyon, 2010; Bauman & Lyon, 2013, Castells, 2010) of the power structures of a Big Data Society, which is what a ‘data ethics of power’ needs to address. That is, a power that is indeed concentrated and engineered by a few power actors, yet also increasingly self-sustained, re-engineered and evolving in (surveillance) cultures (Lyon, 2018) of use, design, governance and imagination, and therefore difficult – but not impossible – to change. I propose that understanding the “liquidity” of power, also means understanding the importance of a holistic data ethical governance approach to the Big Data Society.
CHAPTER 3

Ethical Governance and Sociotechnical Change
MORALITY AND LAW BEGIN when “controversy arises” (Rorty, 1999, p.73). In this way, the philosopher Richard Rorty describes morality as a social and cultural invention that is relational, meaning that it is a type of reflection that springs not from an ethics about the ‘good’ and the ‘bad’ transcendent to a society, but as a response to a disturbance or change of the relations between the components within it. Debating ethics therefore also surfaces when our routines and habits are challenged, and when we question the social construction of what we once knew. Only then “we shall confine ourselves to debating the utility of alternative constructs.” (Rorty, 1999, p. 86) That is, ethical reflections do not come from a predetermined ideal; they emerge out of their relations with other things, things that are transformed and juxtaposed.

In our daily lives, sociotechnical infrastructures are mundane and we take them for granted. They have no visible being as spaces of moral and ethical compromises. Yet, when they break down, malfunction or clash with other legal or moral systems, their embodied moral compromises become visible. This moment where the narrative of one system becomes visible, as we saw in the previous chapter, is also the moment of negotiation that will direct the development of a new system and the transformation of the old.

Conditions for critical ethical reflection can be identified in moments on the macro scale of time of a developing sociotechnical system. These moments, as I argue in this chapter, emerge in between crisis and consolidation in society (Hughes, 1983; Moor, 1985). They are critical moments as they constitute social negotiation and result in the cultural compromises or the “technological momentum” (Hughes, 1983, 1987) that a sociotechnical system needs to evolve. They are also crucial to phases of innovation and development as they constitute the transformation of the sociotechnical system that emerges out of a quest to solve critical problems of that system.

In the late 2010s, BDSTIs and AISTIs were developed and adopted rapidly in public and private sectors and social and ethical implications became apparent at the same speed. As a result, the very design of these sociotechnical systems was increasingly questioned in policymaking and public debate, and new technical designs, business models as well as legal and social requirements emerged. This, among others, also included the
development of a new normality for information computer scientists’ work on BDSTIs and AISTIs, with a new institutionalised framework consisting of standards and laws that directly addressed the data ethical and social implications of their work.

In the previous chapter, I outlined the infrastructural power dynamics of a Big Data Society. In this chapter, I combine an STS and applied computer ethics perspective on technological change to examine the potential role of data ethics in directing this change. I moreover illustrate how the late 2010s represented one such ‘in between’ moment in which the critical sociotechnical problems of the existing technological cultures of BDSTIs and AISTIs became visible and interests and values were explicated and negotiated in ‘data ethics spaces of negotiation’, and consequently, notions of ‘ethical’ and ‘trustworthy’ technology were integrated into the politics, innovation and development of a new system.
SOCIOTECHNICAL CHANGE

What constitutes a change of direction in the way in which a technological system is embedded in society? In this study, I am first and foremost preoccupied with the human shaping of sociotechnical developments. Building on the assumption that developments as such are not just the arbitrary application of technology in society, I argue that we need to consider the different components that may be subjected to human shaping, direction and governance. To do this we also need a conceptual basis for the complexity of factors that constitute sociotechnical change.

As a point of departure, technology is an expression of social practice, created in a dynamic relation between human and nonhuman actors and factors in technological, material, social, economic, political and cultural environments. (Hughes, 1987, 1983; Bijker et al. 1987; Bijker & Law, 1992; Misa, 1988, 1992, 2009, Edwards, 2002; Harvey et al., 2017).

The internet, for example, constitutes a type of applied science and knowledge written down in models, manuals and standards and practiced by engineers and coders, but it is also the result of use in different cultural settings, and it embodies social and legal requirements, political and economic agendas, as well as cultural values and worldviews. That is, technological or scientific processes are not objectively given or representative of natural facts and a natural state of affairs. On the contrary, these very processes and facts are conditioned by their place in history, culture and society and may also be challenged as such.

3.1 Scientific Knowledge and Practice in Paradigm Shifts

As a point of departure, we may consider the very design of a digital data technology a form of applied science, as well as a type of socially shaped and shaping knowledge regime. Accordingly, we can also examine technological change from the perspective of a shift in different paradigms within the foundational knowledge and practice that go into the development of a digital data technology.

Thomas Kuhn (1970) famously examined the historical factors that form paradigm shifts in science, or “scientific revolutions” that shatter the
tradition-bound practices of what is considered normal science (Kuhn, 1970, p. 6). He argued that when big shifts in a scientific field occur, it is not just a question of one theory disproving another, or because a major scientific advancement has been made. Scientific paradigm shifts also involve different ways of seeing the world and doing science accordingly (Kuhn, 1970, p. 16). Scientific revolutions therefore implicate scientific shifts in the very foundational knowledge paradigms of scientific practice, which Kuhn described as particular conceptual, observational and instrumental applications of knowledge (Kuhn, 1970, p. 43). These can be traced in a specific scientific community’s scientific practices as well as in their textbooks and lectures (Kuhn, 1970, p. 43).

Put another way, a paradigm in science represents a way of doing science in a particular scientific community according to a particular social and cultural worldview and priority setting. This means that when a “scientific revolution” occurs it is not only the science that changes. According to Kuhn, scientific paradigm shifts do indeed often involve the rejection of one scientific theory in favour of another, but there is also a fundamental change in what a scientific community counts as a problem worth solving and a change in their standards and methodologies; that is, the “scientific imagination” of the field and the kind of educational and instrumental environment that the science is applied in. Accordingly, these paradigm shifts are major and revolutionary in that they uproot everything that is considered normal in the field, creating not just new theories but also new practices, instruments and objectives: “Such changes, together with the controversies that almost always accompany them, are the defining characteristics of scientific revolutions.” (Kuhn, 1970, p. 6)

In this manner, paradigm shifts in science were described by Kuhn as revolutionary earth-shattering changes in the way we conduct, develop, and receive a science, transforming normality by changing the very foundations of what we assume to be its purpose, shape and direction. To exemplify this, Kuhn quoted Einstein’s autobiographic description of the early moment of the scientific revolution of physics he spurred: “It was as if the ground had been pulled out from under one, with no firm foundation to be seen anywhere, upon which one could have built.” (Kuhn, 1970, p. 83). This also means that scientific change necessitates not only the acceptance of a new normal in how a scientific field is practiced and conceived of, but also in how it is governed in society, and accordingly it requires new scientific practices,
habits and standards.

3.2 The Complexity of Change: Conflicts, Negotiation and Compromise

Considering technological change in terms of paradigm shifts is helpful when examining how the institutionalised standards as well as the foundational worldviews and knowledge frameworks for the work of the people who build technological systems transform and technological practices change. However, an analysis here that focuses only on the transformation of technological components would be incomplete. Technology is shaping society, but it is also adapting according to societal factors. A diverse and complex set of social, political, economic, cultural and technological factors come together when large sociotechnical infrastructures, such as BDSTIs and AISTIs, transform and evolve in society. Understanding technological change means discerning the complexity of these components in terms of their negotiations, and thus, the compromises that they embody. As Bijker and Law stated:

“Technologies always embody compromise. Politics, economics, theories of the strength of materials, notions about what is beautiful or worthwhile, professional preferences, prejudices and skills, design tools, available raw materials, theories about the behavior of the natural environment – all of these are thrown into the melting pot whenever an artefact is designed or built.” (Bijker & Law, 1992, p. 3).

Investigating the diverse factors that form the shape of a sociotechnical system also empowers human actors to direct its evolution. One way of doing this is to explore the different interests invested in socio-technical change as well as their conflicts and negotiation. This also implies an examination of the moments of controversy and conflict in which core problems of technological systems are identified; solutions to problems, successes or failures of systems are negotiated; and priorities and goals for the evolution of the sociotechnical system are set (Hughes, 1983, Hughes, 1987, Misa, 1992). Following these controversial moments, a sociotechnical system is stabilized, generally accepted, and consolidates in society. It becomes, so to speak, the state of affairs. It is this focus on the very conditions for technological change that makes the trajectory of a technology’s development and societal adoption manageable by humans. As Francesco Lapenta (2017) states, the future is “…not arbitrary but the product
of a complex series of decisions and actors that can potentially give shape to a number of differently possible, probable, or desirable future scenarios.” (Lapenta, 2017, p.154). That is, the often conflictual negotiations for technological development of the present (as the ones we for example saw in the public debates on BDSTIs and AISTIs in the 2010s) must also be thought of as reflective choices about the future.

Here, I argue that ethical reflection about the ethical compromises and the trade-offs between different values and interests we make should be a core component of these moments of controversy.

### 3.3 The Four Phases of Sociotechnical Change

One way of understanding the transformation of sociotechnical systems is to study how patterns in their diverse components evolve over time. Grasping developments on a macro scale allows us to intervene in moments open to critical intervention to shape the direction of evolving sociotechnical systems.

A key theory of the change of large sociotechnical systems is Thomas P. Hughes’ (1983) analysis of the phases of the development and expansion of the world’s electric power systems between 1880 and 1930. With a description of the specific complex economic, political, social and scientific components of the different phases of this development, he also illustrated more generally how technological systems evolve in patterns over time in constant dialogue with the interests embedded in their environments.

Hughes argued that although larger sociotechnological systems are instituted in different places and reach the different phases of development at different times, they evolve and expand according to a model pattern consisting of phases that are characterised by their dominant activities: invention, development and innovation, transfer, growth, competition, and then finally consolidation (Hughes, 1983, 1987).

**Invention and development:** The first phase is characterised by inventors and entrepreneurs that are the key drivers for the invention and the initial development of the system. Here, managers, engineers and financiers are involved secondarily (Hughes, 1983, p. 14).

**Transfer:** In the second phase the focus moves to the process of
transferring the technology from one region and society to another and equally the dominant agents involved in this phase change to include, in addition to the entrepreneurs and inventors, the financiers and organisers of enterprises as key actors (Hughes, 1983, p.14).

**Growth:** In the third phase a range of actors, entrepreneurs, inventors, engineers and others, dedicate their efforts on correcting and finding solutions to what Hughes refers to as their “reverse salients” that are formulated as critical problems that prevent the system from growing (Hughes, 1983, p.14).

**Momentum, competition and consolidation:** A large sociotechnical system requires a momentum with “mass, velocity and direction” which is created by the different interests invested in the system in the fourth phase of societal consolidation (Hughes, 1983, p.15).

These descriptions of the components of particularly the third and fourth phases of the sociotechnical development of larger technological systems are helpful when describing the development of BDSTIs and AISTIs in the late 2010s. To start with, Hughes refers to “a battle of the systems” in which an old and a new system exist simultaneously in a relationship of “dialectical tension” (Hughes, 1983, p. 79). He described this third phase as a moment of conflict and resolution, not only among engineers but also in politics and law (Hughes, 1983, p. 107). In these moments of conflict, critical problems are exposed, different interests are negotiated, and then they are finally gathered around solutions to direct the evolution of the systems. In the growth phase just before consolidation, “reverse salients” are formulated as critical problems.

Hughes described a “reverse salient” as a component of an expanding system that “...does not march along harmoniously with other components. As the system evolves toward a goal, some components fall behind or out of line” (Hughes, 1983, p.79). Accordingly, in this phase there is also an intense focus on problem identification, and solutions are proposed and negotiated by various actors. The new system, or the transformation of the old system, evolves out of the very problems identified and solved in this phase. A “reverse salient” may be technical problems, but they may also be financial or organisational (Hughes, 1987, p. 74), and once identified, a group of
“problem solvers” from inventors, engineers, and managers to financiers and legal experts takes over to create solutions for them (Hughes, 1987, p. 74).

“Reverse salients” may arise from inside the very systems or from their immediate environment, but crucially Hughes argued that they are bound by time and place (Hughes, 1983, p. 80). In other words, the critical problems of the systems are not just resolved as technical problems, for example, with agreement on technical standards with systems requirements, but they are in dialogue with political and historical factors. In contrast to Kuhn, Hughes did not describe the phase of conflict and resolution as necessarily a revolutionary one; that is, as one in which one paradigm is replaced by another incompatible with the first one. The systems change in “synthesis” and in a combination of “coupling and merging” between the old and the new systems, which gradually evolve over decades and on different levels from the technical to the institutional with invested interests gradually transferring from one system to the other (Hughes, 1983, p.121). Only in cases where a “reverse salient” cannot be resolved within the system the solution need to be found in the development of a radically new system (Hughes, 1987, p. 75).

Now, applying Hughes description of the patterns of particularly the third and the fourth phases of sociotechnical change to the development of BDSTIs and AISTIs in the late 2010s, there are recognizably similar patterns. In the early 2010s, a global big data digital infrastructure connected different regions of the world, cutting across jurisdictions and thus challenging their traditionally territorial scope (see Hasselbalch, 2010). Most profoundly, the legal right to privacy and personal data protection was challenged by this new technologically enabled interjurisdictional space in which different levels of protection and safeguards were required and implemented in the design of data technologies and systems. Clashes between different regions’ legal frameworks for protecting privacy and for protecting business or state interests in data emerged, and as I describe in the first article of this study (Hasselbalch, 2019), various privacy-by-design technologies and systems were initially proposed as solutions to the critical problems of the ‘old system’.

In the mid-2010s, critical problems concerning privacy rights and personal data protection became, as described in the previous chapter, particularly prevalent following Edward Snowden’s revelations of mass surveillance and major data hacks of online services, such as the social
networking service Snapchat in 2013 or the infidelity site Ashley Madison in 2015. Such critical problems were revealed and identified by activists, whistle blowers and journalists and picked up by engineers and policymakers from different regions of the world, who would propose, impose and design technical and legal solutions (Hasselbalch & Tranberg, 2016).

Here, one can consider Hughes’ depiction of “reverse salients” as components of a system that fall out of line or are disharmonious with other components of the system, and therefore freeze the system’s consolidation in society (Hughes, 1983, p.79). The critical problems of BDSTIs, and also of AISTIs, that surfaced at this time, particularly regarding the protection of personal data and online privacy, did indeed halt the ongoing consolidation of BDSTIs, causing first and foremost a battle between different regional legal governance approaches to the technical development and business conduct behind these systems. These were the “reverse salients” of BDSTIs restricting a global big data system in growth and consolidation, and importantly limiting its momentum in society with conflicts between business interests, citizen interests, state agency interests, and political and regional interests.

For example, a series of big data social networking services developed predominantly in one U.S. area, namely Silicon Valley, had in the 2000s cut across the globe practically unnoticed by legislators, and therefore in the 2010s were already silently consolidating in European people’s social and private lives (e.g., 44% of Europeans said they never used social networking services in 2011 [Eurobarometer 76], while only 28% said this in Autumn 2019 [Eurobarometer 92]). These services represented two different conflicting goals: to connect and facilitate information exchange, communication and the social life of people, and also to provide companies with new means of micro targeting customers with marketing, as we saw in a previous chapter of this study. On these grounds, a battle and critical space of negotiation emerged with proponents of a big data business model and a new emerging privacy-by-design business and activist movement (Hasselbalch & Tranberg, 2016). It, for example, enticed the tougher data protection legal provisions of the European General Data Protection legal reform, which was negotiated between 2012 and 2016.

All in all, the mid- and late-2010s were characterised by a “battle of systems”, the moment of conflict in which technical, legal, cultural and social
components of an old and a new system existed simultaneously in a relationship of “dialectical tension” (Hughes, 1983, p. 79). The “reverse salients” of BDSTIs and their AISTI evolution were identified, in politics and public opinion in particular, as ethical and social critical problems of the existing systems’ data handling and design (as I illustrate in the three articles of this study). That is; “reverse salients” were approached as sociotechnical problems. Accordingly, at the end of the 2010s, not only were engineers proposing and negotiating solutions, but also an increasing number of often new types of scientists and experts, combining humanistic studies, social science and philosophy with data science, were participating in negotiations to identify critical problems and propose, among others, applied ethics solutions. These solutions were shaped as responses to the ethical challenges specific to BDSTIs and AISTIs. Therefore, this moment also took form as a negotiation of the societal and ethical values that were to shape the direction of the BDSTIs and AISTIs’ technological momentum; that is, their “…mass, velocity, and direction” (Hughes, 1983, p.15).

3.4 Ethics in the Policy Vacuums of Sociotechnical Change

In 1985, James H. Moor predicted a computer revolution of society in which ethical reflection and value negotiation is brought about. His view was specifically fixed on the social effects of the introduction of computers and ethical reflection as a direct response to the specific characteristics of computerised environments. However, in combination with Hughes’ depiction of the more general phases of sociotechnical change, it is relevant here to include Moor’s view on the role of ethics when the norms and values we take for granted are challenged due to a sociotechnical transformation, such as the one introduced to society by computer technologies.

Moor proposed that the computer revolution would occur in two stages, marked by the questions we ask. In the first “Introduction Stage”, we will ask functional questions: How well do particular technologies function for their purpose? In the second “Permeation Stage”, when institutions and activities are transformed, we will ask, as we also saw in the late 2010s, questions regarding the nature and value of things (Moor, 1985, p. 271). That is, ethical reflection emerges in situations where computers alter situations and clash with existing policies.
Moor also foresaw that the adoption of computers in society would “...leave us with policy and conceptual vacuums...” (Moor, 1985, p. 272). These vacuums present core critical problems and challenges, almost like Hughes’ “reverse salients”, but rather than problems specific to the sociotechnical systems, they present challenges to the specific social environments in which they have been implemented and policies have been established. In this way, they reveal conceptual muddles and uncertainties and we are presented with new choices of action (Moor, 1985, p. 266). It is exactly due to this juxtaposition between what we once knew and what we now do not know that we are forced to reflectively consider what we find valuable. That is, the very clash between the technological system (the computer) and existing policy frameworks that we have previously taken for granted will force us to “...discover and make explicit what our value preferences are.” (Moor, 1985, p. 267).

I propose the use of Hughes’ theory of sociotechnical change in combination with Moor’s depiction of policy vacuums to understand the role of data ethics in governance and policymaking. If we for example interpret the data ethics public policy initiatives launched in Europe in the second half of the 2010s (that I describe in the first article of this study) in this perspective, we see that they are actually not solutions to the ethical problems that they explicitly addressed; rather, they are ‘spaces of negotiation’ in which more general cultural values in conflict are negotiated. These were the policy initiatives; however, several other critical ‘spaces of negotiations’ emerged in the late 2010s that were particularly critical of the powers of BDSTIs and AISTIs and their main stakeholders, such as the Google employee walkouts to protest against the treatment of women in the company in 2018, or the UK student demonstrations against an automated A-level grading system in 2020.

I propose that we acknowledge the ethical compromises we make in these moments of value negotiation as valid components of the governance efforts invested in sociotechnical change. They constitute the very cultural compromises shaping the technological momentum.
GOVERNANCE

Through a sociotechnical analysis, I seek to delineate the combination of not only technical but also social, cultural and economic components that constitute the shape of a technological development. I do this to entice an apprehension of the factors that in particular policymakers need to address to guide this change with a human (-centric) approach. As Bijker and Law (1992) put it, technologies do not represent their own inner logic; they come out of a range of heterogeneous factors, but nevertheless they are also shaped, even “pressed” into a certain form, that “might have been otherwise” (Bijker & Law, 1992, p. 3). This is, as I have argued before, an essential view on technological development and change, such as the evolution of BDSTIs and AISTIs, as it empowers human governance efforts and also prevents a single-sided analysis of the components of BDSTI and AISTI development that we target with governance initiatives.

For example, if we want to steer the interests invested in a BDSTI’s technological momentum within a specific value framework, we will not succeed by, for instance, investing only in the development of ‘XX by design’ technology. Neither will we effect change in this manner by raising the awareness of consumers alone, nor by regulatory requirements only, or with the creation of new systems requirement standards alone. We require a distributed governance approach where each of these activities is considered a component of a totality that addresses the ethical implications of complex social, cultural and political environments.

Furthermore, the ‘we’ that is doing the shaping is not an actor that can be identified in just one place and as just one actor (Harvey et al., 2017; Hofman et al., 2017; Mueller, 2010; Brousseau & Marzouki, 2012; Epstein et al., 2016). Indeed, technological systems are, like legislators, actors of governance with active socially ordering and governing powers, but so are users, engineers and designers of technological systems (DeNardis, 2012; Winner, 1986). As we will see in the following discussion of internet governance, the ‘governing’ of sociotechnical change is a complex heterogenous process.

3.5 Internet Governance

There can be no infrastructure without some type of governance. Shared
frameworks of rule-making, ordering and collective action will always be core to a functioning sociotechnical infrastructure (Star & Bowker, 2006, Bowker et al., 2010). On a very basic technical level, without these shared constructed frameworks, the technical components of the system do not interact and the system breaks down or its development is halted. The same counts for legal frameworks, such as laws on how to protect and share data and to protect privacy rights of individuals, as I have described before in this study. They need shared frameworks to function on a very basic level of application. That is, while the negotiation between different values, conflicts of interests and battles of systems in critical moments of a system’s development may represent the uncertainty of shared governance agendas, or what Moor refers to as “policy vacuums”, a well-functioning infrastructure requires a level of agreement to work. This is also what Star and Bowker (2006) refers to as “handshakes” between the different components. A working infrastructure will therefore also always represent a compromise or the domination of one standard (legal, cultural, social) over others.

The internet is an example of a large-scale information infrastructure that obviously requires institutionally shared global governance technical – as well as policy and legal – standards to operate efficiently. All the same, in the early 1990s the World Wide Web came into being with a cyber-libertarian imagining of an independent public sphere in which citizens were set free from oppressive state governance by the very decentralised and ungovernable information architecture of the digital network (Mueller, 2010, p. 2). From this, a conception of a different type of bottom-up, people-centred and ethics-based form of governance emerged. As John Perry Barlow famously wrote in his declaration of the independence of the internet in 1996:

"We believe that from ethics, enlightened self-interest, and the commonwealth, our governance will emerge. Our identities may be distributed across many of your jurisdictions. The only law that all our constituent cultures would generally recognize is the Golden Rule. We hope we will be able to build our particular solutions on that basis. But we cannot accept the solutions you are attempting to impose.” (Barlow, 1996).

Imaginings like this of technological liberation and freedom from institutional governance lived long into the formative years of the global sociotechnical information infrastructure of the internet (Mueller, 2010, p.
In the 2000s, however, political battles and negotiations between traditional governments and intergovernmental institutions on the global scene had reached new levels with several regulations and policy initiatives introduced specifically dedicated to the governance of a new internet-based sociotechnical sphere (Brossau & Marzouki, 2012). The internet was not just a free zone for the ‘ungoverned’ emancipation of the individual; it was still constituted with a form of governance, although governance of this new sociotechnical public sphere was also not a state-only activity, and increasingly neither was it recognised as such in the official policy sphere.

New actors of governance were emerging and positioning themselves in internet governance policy debates—the engineers and businesses, internet users and their communities (DeNardis, 2012; Harvey et al., 2017; Hofman et al., 2017; Mueller, 2010; Brousseau, E., & Marzouki, 2012; Epstein et al., 2016). Most significantly, large corporations were designing, but also setting the rules and codes of conduct for their online platforms (Aguerre, 2016; Belli & Zingales, 2017, Wagner et al., 2019; Franklin, 2019; Jørgensen, 2019). That being so, in the early 2000s, multistakeholder governance institutions and initiatives were introduced in institutionalised form. By way of illustration, the UN Internet Governance Forum (IGF) was formed during the initial World Summit of Information (WSIS) processes in 2003–2005, in which the identification of and solutions to problems went far beyond the mere technical design of the internet due to the involvement of civil society and technical community stakeholders (Brosseau & Marzouki, 2012).

Many internet governance studies have focused on the dynamics of the governance of the internet, as well as ‘how’ to govern a disruptive global and interjurisdictional sociotechnical information infrastructure (DeNardis, 2012; Harvey et al., 2017; Hofman et al, 2017; Mueller, 2010; Brousseau, E., & Marzouki, 2012; Epstein et al., 2016). As a result, most internet governance scholars move from the perception that the very technological architecture of the internet has brought about new governance models and as such disrupted traditional centralised forms of governance. Milton Mueller (2010), for example, describes how the internet as a technological architecture imposes on nation state governance in different ways. The very cross-border communicational technical architecture of the internet means that attempts to impose jurisdictional additional architectures require extra efforts (Mueller, 2010, p. 4). In addition, the architecture of massive information generation, collection and retrieval enables massive-scale
communication, which traditional governments have difficulties in responding to, and which also transforms their governmental processes (Mueller, 2010, p. 4). Moreover, the decentralised and distributed internet architecture distributes control, and new transnational institutions (such as ICANN) form a new type of power centre for key decisions (Mueller, 2010, p. 4). Ultimately, the internet transforms “the polity” with new types of collaboration, organisation and mobilisation across borders by converging media and creating new types of communication that lowers the cost and empowers group action (Mueller, 2010, p. 5). Based on these observations, Mueller uses the term ‘governance’ rather than ‘government’ to shift the focus from traditional forms of centralised rulemaking and social ordering steered by nation states. The global sociotechnical infrastructure of the internet has indeed disrupted nation state governance; however, this does not imply that it is not still directed and shaped. All this means is that governance is “less hierarchical and authoritative” (Mueller, 2010, p. 9).

Internet governance scholars have been particularly preoccupied with the initial official first attempts to negotiate on an institutional level a shared global governance approach to the internet and what was also increasingly delineated in global policymaking as the Information Society in the WSIS process initiated in the early 2000s and the following IGF hosted by different countries worldwide every year (Mueller, 2010; Epstein, 2013; Bygrave & Bing, 2009; Flyverbom, 2011; Brosseau & Marzouki, 2012 ). Increasingly, an STS-informed approach has also been used to analyse a more complex set of governance actors and models that emerged in the sociotechnical formation of the internet in society (Epstein et al., 2016). Many different governance components are acknowledged to shape the direction of a sociotechnical system (laws, cultural norms and habits, education, manuals for engineering practices, standards, funding schemes, and codes of conduct).

Epstein et al. (2016) delineates the key components of an STS-informed approach to internet governance as based on the following foundational ideas: Firstly, there is a “plurality” of modes of governance that are also taking place and being enforced in a variety of fora and according to a diverse set of “normative systems” from law and technology to social practices. Then, the technical infrastructure has ‘nonhuman’ agency that not only orders the social but also controls it. Moreover, it is not only the official and extravagant actions (such as regulations and political agendas) of
humans that ‘govern’; it is also the invisible “mundane practices” of humans that shape “design, regulation and use of technology”. Importantly, a key focus is “controversies as structuring and performative processes” where different actors of interests are juxtaposed and negotiated, exposing their different notions of governance. Finally, the notion of multistakeholderism is brought forward by an STS-informed approach when acknowledging the many actors that participate in ‘doing internet governance’, and specifically in the role of private actors (from users to industries) in the context of decision-making and governance of the internet as such. (Epstein et al., 2016, p. 6-7).

The term “reflexive coordination” is introduced by Jeanette Hoffmann et al. (2017) in an attempt to embrace these heterogenous components of internet governance:

“Our approach on governance proposes a fundamental shift in perspective: instead of gradually extending a regulatory perspective beyond nation-states, public decision-making and formal policy instruments, we suggest studying Internet governance as a continuous heterogeneous process of ordering without a clear beginning and endpoint.” (Hoffman et al., 2017, p. 1412).

This is a type of governance that takes place ‘in-between’ the top-down, intentional steering of states and the heterogeneous disruptive and less organised ordering activities of dispersed (new and traditional) actors of internet governance (the state, engineers, users, citizens, scientists, and technological artefacts), considering their intertwined intentional and unintentional “multiple orders” (Hoffman et al., 2017, p. 1410). By focusing on the way in which different actors are coordinated and interrelated, the complexity and diversity of different actors of governance are brought to light. Their coordination activities might be simple and ordinary and – on the face of it – unexciting. Nevertheless, they do create a type of social order (Hoffman et al., 2017, p. 1413). This type of routine and habitual coordination of order can only be steered reflectively (“reflective coordination”) in critical moments when different norms, assumptions and understandings of situations clash (Hoffman et al., 2017, p. 1415).

I wish to use these reflections on governance from a field of internet governance studies to explore the role of data ethics in the ‘governance’ of the sociotechnical change of BDSTIs and AISTIs. They are descriptions of new legitimate modes of governance emerging as a response to, but also
reinforced by, the specific architecture of the internet. It is a type of steering of sociotechnical change that is institutionally engineered as well as nonengineered in cultural practices of, for instance, engineers and users. Governance can also be understood here as ‘open-ended’ in the sense that it does not have a predefined start and end point or solution that we can steer towards. A law reform, for example, is not only a clearly delineated negotiation process that takes off with a proposal and ends with the adoption of a new law; it, among others, for example also consists of follow-up evaluation mechanisms (Brøgger, 2018).

Here, I also want to combine the critical moments in which simple coordination activities transform into “reflexive governance” (Hofmann et al., 2017) with Hughes’ moments of conflict in a technological system’s development, where “reverse salients” are identified as critical problems of the system and dialectical tension between different systems occurs (1983,1987). Most importantly, I want to relate them to the kind of ‘ethical situations’ that Moor describes, where the problem-solving and negotiating actors involved in the computer revolution of society are more focused on conceptual activities and explication of the nature and value of things due to “conceptual muddles” and uncertainties of policy frameworks (Moor, 1985, p. 266). That is, they are the ethically reflective moments that emerge out of relations, as Rorty describes, when “controversy arises” (Rorty, 1999, p.73), and thus, they are the essence of the type of ‘data ethical governance’ I want to propose as the governance component of a ‘data ethics of power’. This is when standard world views, norms, and data practices and cultures clash and force out a particular type of reflection of the social construction of our assumptions; and accordingly, an ethical reflection on alternative data policies and practices emerge (Moor, 1985).

These moments are also moments of values-based governance, characterised by the inclusion of multiple actors with an interest in the data of BDSTIs and AISTIs, a consideration of intertwined intentional and unintentional “multiple orders”, and reflective coordination between them (Hoffman et al., 2017). It is when established norms and values in data engineering and design processes as well as use are challenged, and when data ethical problems are identified and policies, strategies and solutions are specifically formulated to address them. This is also the way in which Rainey and Goujon (2011) describe “ethical governance”, as a reflexive rather than top-down approach that takes into account the conditions for ethical
reflection in particular:

“What’s required is an approach that can offer first criteria of evaluation and second a more interesting way to address the conditions not only for an ethical reflexivity, but also for determining the conditions of construction of ethical issues, of ethical norms, and the conditions for their adoption and implementation” (Rainey & Goujon, 2011, p. 54).

3.6 Ethical Governance

At Google’s annual developers’ conference in 2017, CEO Sundar Pichai reiterated the company’s “AI-first” mission to make machine learning an umbrella integrated into all Google platforms to enhance all services from video, search and email to mobile. In a FastCompany interview, Pichai remembers this approach to put AI first in all Google products as a moment of existential revelation:

“This thing was going to scale up and maybe reveal the way the universe works (...) This will be the most important thing we work on as humanity.” (Brooker, 17th September, 2019).

In the late 2010s, this type of all-embracing and dedicated AI approach was not unique to Google. It was evident in BDSTI developments in general. BDSTIs were developed and adopted with a sense of urgency similar to the urgency of the big data imagination of the 1990s. In this way, BDSTIs were also gradually transforming into AISTIs with advanced technical components designed to sense their environment in real time, learning and evolving autonomously. AISTIs had social components; facilitated and increasingly constituted public and private spheres; and were already in part institutionalised in systems requirements standards for IT practices and regulatory frameworks for data protection. However, concerns regarding the ethical implications of autonomous AI systems were increasingly also raised within engineering communities towards the end of the 2010s. Thus, in 2019, at Google’s annual developers’ conference Jen Gennai, Google’s head of responsible innovation, told the crowd of developers: “We’ve identified four areas that are our red lines, technologies that we will not pursue. We will not build or deploy weapons. We will also not deploy technologies that we feel violate international human rights.” (quoted by

---

2 See Sundar Pinchai’s talk at Google’s 2017 annual conference: https://events.google.com/io2017/
Brooker, 17th September, 2019). The crowd of developers were not convinced about Google’s more contained approach to AI though. As one developer that used to work for Google but had quit in protest said to the FastCompany journalist present at the event: “I was pushing for: What are things you would never do? I never got clarification.” (quoted by Brooker, 17th September, 2019)

Any attempt to govern the direction of sociotechnical systems, such as the evolution of BDSTIs into AISTIs, would have to take into account a complex intertwined network of relations, which includes the different worldviews and imaginations driving forward particular scientific and technical developments. Specifically, as I have argued, the late 2010s presented a moment of conflict, a “battle of systems” (Hughes, 1983), which I argue also constituted an uprooting of what had previously been taken for granted in big data technology practice. As such, the development of BDSTIs and AISTIs was invested with big data and AI-first ‘mindsets’, but these were increasingly also challenged – as I illustrate in the following chapters and describe in the articles of this study – by competing values-based approaches, such as ‘privacy by design’, ‘ethical design’, and conceptions of ‘human-centric’ and ‘trustworthy AI’.

In this study, I focus on the role of ethics in critical moments such as this; how we in these moments negotiate and explicate foundational values and interests, and then within this crisis and space of negotiation create new policies and directions for sociotechnical systems. Ethics, I argue, play a crucial role in governance as an explicit reflection of values that serve a human interest in spaces of negotiation between different interests. As described, the technological momentum required for a large sociotechnical system to consolidate in society is not just an arbitrary composition of social, economic and cultural factors mixed together by an inexplicable will of nature; it has a shape that guides the direction, values, knowledge, resources and skills that form the technological architecture of the system—its governance, adoption and reception in society. At times, this shape is more explicitly cultural and values-oriented than others (Moor, 1985).

I wish to propose that perhaps this cultural shape, the reflective ethical evaluation and value-orientation, is a moment we may consider not just a momentary critical response to critical problems but also a valid component of governance. I also argue that in the 2010s, ‘data ethical governance’ was indeed recognized as such in public policymaking, where ‘data ethics policy
initiatives’ were increasingly accepted as components of institutionalised forms of governance (as described in the first article of the study).

Winfield and Jirotka (2018) also use the term “ethical governance” to present a case for “… a more inclusive, transparent and agile form of governance for robotics and artificial intelligence (AI) in order to build and maintain public trust and to ensure that such systems are developed for the public benefit.” (Winfield & Jirotka, 2018, p.1). “Ethical governance”, they argue, goes beyond just good and effective governance. It is “…a set of processes, procedures, cultures and values designed to ensure the highest standards of behaviour.” (Winfield & Jirotka, 2018, p. 2). Governing the development of robotics and AI with an ethical framework, they therefore argue, requires a diverse set of approaches, from those at the level of individual systems and application domains to those at an institutional level (Winfield & Jirotka, 2018, p. 2). Thus, they also argue that “while there is no shortage of sound ethical principles in robotics and AI, there is little evidence that those principles have yet translated into practice, i.e. effective and transparent ethical governance.” (Winfield & Jirotka, 2018, p. 9). Along these lines, I argue that “ethical governance” concerns not only the negotiation of foundational ethical values per se, but also comprises the very practices and processes that produce ethical reflection, so to speak.

3.7 Data Ethical Governance

Expanding on the previous discussion of internet governance and ethical governance, I here want to examine the type of ethical governance that addresses the complexity of the Big Data Society in particular. Drawing on the discussion of internet governance and ethical governance as a multi-actor, reflexive, open-ended (Hoffman et al., 2017) and agile process designed to ensure the highest standards of behaviour (Winfield and Jirotka, 2018), a range of real-life activities occurred in the 2010s with a somewhat shared aim of inducing human-centric data cultures (see below). These activities can also be described as the type of critical applied ethics that I in this study consider the practical implementation of a ‘data ethics of power’.

Whereas applied ethics is traditionally considered a branch of a normative moral philosophy framework, I take an approach to applied ethics that is informed by pragmatism. Pragmatism parts from normative moral philosophy in that it first and foremost emphasises applied ethics as a practice and reflective process rather than an implementation of an applied
normative framework. That is, an ethical theory or approach can only be justified in practice. Philosopher of law and politics Andrew Altman described pragmatism as an applied ethics that involves a thorough “contextualist view of justification” (Altman, 1983, p. 232). As such, there is no neutral ethical starting point (Altman, 1983, p. 232). In fact, a fundamental idea is that any ethical assumption can be challenged in context:

“Perhaps there are universally valid ethical principles (...) The pragmatist merely insists that the justification of any claim to the universal validity of some principle can be made only in the context relative way...” (Altman, 1983, p. 232).

An example of the two different applied ethics approaches can be found in studies of ethics guidelines and principles from the late 2010s. During this period, a myriad of normative ethics guidelines and principles were created worldwide to address the ethical implications of AI in particular by various stakeholder interest groups (Jobin et al., 2019; Fjeld et al., 2019). One response was to create a common normative framework based on universal principles that were traced in thematic convergences between the various documents (Winfield & Jirotka 2018; Floridi et al., 2018).

However, a pragmatist applied ethics approach would take a different route. Here, each guideline would be considered in its relative context in terms of its unique point of reference, its ‘non-neutral’ ethical point of departure, and thus importantly in terms of the societal power dynamics in which it is positioned within distinct cultural, social and legal frameworks. I illustrate this proposition in the second article of this study when analysing the cultural framing of the EU’s AI agenda.

Critically, Altman defines a pragmatist applied ethics as any activities that seek to justify an ethical theory in practice, and concludes that “applied ethics would include any activity in which an individual was consciously testing some normative principle or theory.” (Altman, 1983, p. 233). Altman here uses as the example of Karl Marx’s historical-dialectical materialism, which he saw only as justified if practiced in a social revolution (Altman, 1983, p. 232). He also uses the more grounded example of the Philosopher John Dewey’s Laboratory School of Chicago created to test his theories about experience, the mind and society (Altman, 1983, p. 233).

The following are key subfields and examples from the 2010s of what I, in a pragmatist framework, consider applied data ethics or ‘data ethical
governance’ in terms of their constructive approaches to the power of sociotechnical data infrastructures. I do not (and cannot) provide exhaustive lists of the activities in these subfields, but simply provide a few examples within each.

- **Data Ethics Initiatives**

In the late 2010s, a number of initiatives emerged in the policy, business and technology fields with explicit reference to data ethics. In the first article of this study, I describe the public policy data ethics initiatives in Europe and argue that these emerged as ‘spaces of negotiation’ created out of controversy regarding a presumed neutrality of BDSTIs. Specifically, they often created explicit conflicts between different data cultures and the ethical implications of the dominant data cultures of data design and practice. These appeared alongside several civil society, academic and technology initiatives in which data ethical implications were framed and sought to be resolved as an issue of a growing data asymmetry between big data institutions and citizens in the very design of data technologies. As an illustration, the conceptual framework of the “Personal Data Store Movement” (Hasselbalch & Tranberg, 27 September 2016) was described by the nonprofit association MyData Global Movement as one in which “[i]ndividuals are empowered actors, not passive targets, in the management of their personal lives both online and offline – they have the right and practical means to manage their data and privacy” (Poikola et. Al., 2018). Here, the emphasis was on moving beyond mere legal data protection compliance to implement values and ethical principles such as transparency, accountability and privacy by design (Hasselbalch & Tranberg, 2016). In particular, mitigation of the ethical implications was sought with values-based approaches to the design of technology; for example, engineering standards such as those of one of the world’s largest engineers’ organisations, IEEE’s, P7000s series of ethics, and AI standards that strived towards

---

8 It must be emphasized here that these examples are limited in that they primarily draw attention to my own context. A further elaboration of an applied ‘data ethics of power’ would benefit greatly from a non-Western focus.
developing ethics by design standards and guiding principles for the development of AI.9

- **The Privacy Civil Society Movement**
A crucial component of the data ethics momentum of the 2010s were the many consumer and citizen awareness initiatives launched in different civil society contexts with reference to the power asymmetries online between citizens, states and the private industry. In the 1990s, the privacy movement was dedicated to the development of privacy-enhancing technologies (PET) in more technically savvy contexts, with among others the introduction of the TOR anonymity software and the TOR project and movement (Hasselbalch & Tranberg, 2016, p. 85). However, in the early 2010s the privacy movement was starting to take a more popular form with organisations, such as the UK-based Privacy International and US-based Electronic Frontier Foundation (EFF), dedicating campaigns to citizen online privacy awareness in regards to state surveillance practices in particular. The crypto party movement initiated by the Australian journalist Asher Wolf in 2012 resulted in a range of self-organised crypto parties worldwide, which citizens could attend to learn how to protect their privacy and anonymity online. Increasingly the popular privacy movement was also taking into account the big data practices of the private industry, offering “digital self-defence” tools and alternatives to the big data technology industry giants consumer services (Tranberg & Heuer, 2013; Hasselbalch & Tranberg, 2016; Veliz, 2020).

- **Ethics by Design and Investigations of ‘Data Systems’**
A specific applied ethics focus on technology and design was spelled out in numerous ‘ethics by design’ activities. ‘Ethics by design’ is a term used to address the design and design practices of a technology. The understanding is that the very design of a technology may reinforce, alter or even produce values and power relations. Phrased differently, the presumption is that technology has “politics” or embedded “arrangements of power and authority” (Winner, 1980, p. 123). This approach has been used specifically in the context of bestowing ethical values in the design of autonomous

---

9 See Ethics in Action, P7000s standards. https://ethicsinaction.ieee.org/p7000/
agents (i.e., AI) (Dignum et al., 2018). It is rooted in the value sensitive design (VSD) approach, which is generally applied in the field of computer technology development and which I describe in more detail in Chapter 4.

Here, it is important to mention the “privacy by design” approach, developed originally by Anne Cavoukian (Cavoukian, 2009), as it specifically focuses on organisational and design practices that seek to embed ‘privacy’ as a value in the data design of a technology (I return to the ‘ethics by design’ and VSD approach in the second part of this study). In addition to the design focus of a VSD or ‘ethics by design’ approach, we may also include here what the philosophy and technology scholar Philip Brey describes as a “Disclosive Computer Ethics”, which seeks to identify and reveal ethical implications in opaque information technologies (Brey, 2000, p. 12). A range of case studies of specific data processing software and their use are worth mentioning. Examples are the “Machine Bias” study (Angwin et al., 2016), which exposed discrimination embedded in data processing software used in U.S. defence systems; Safyia Umoja Noble’s (2018) investigation of Google’s discriminatory search algorithms; or Cathy O’Neil’s (2016) analysis of the social implications of the math behind big data decision-making in everything from obtaining insurance and credit to getting and holding a job.

• The Law
A range of legal studies have critically assessed the legal framework for big data sociotechnical development, privacy (Solove, 2006; Cohen, 2012 ), human governance, and the rule of law and human rights implementation in autonomous data-based systems, as well as AI and robotics (Pasquale, 2015, 2020; Smuha, 2020; Nemitz, 2018; Latonero, 2018; Hildebrant, 2016). In the late 2010s a number of legal studies focused on the European data protection legal framework GDPR (for example, Wachter et al., 2017; Zarsky, 2017; Wachter 2019). There were various emphases on specific challenges in law in regards to children’s personal data (Hof, 2019), for example, or in the context of smart toys (Keymolen & Hof, 2019). Here, it is also relevant to mention work on a legal framework for ‘data trusts’ presented by Sylvie Delacroix and Neil D. Lawrence (2019). They consider the development of a plurality of ‘data trusts’ that individuals can choose between – an empowering alternative to what they refer to as a “one size fits all’ approach to data governance”, since that will allow ”data subjects to choose a Trust that reflects their aspirations, and to switch Trusts when needed” (Delacroix &
Lawrence, 2019, p. 236).

- The Discourse
Critical data studies have deconstructed the cultural narratives of dominant data cultures of institutions, industries and communities that design and build the systems in which big data is processed and analysed (from social networking services to AI agents; Bowker & Star, 2000; Kitchin & Lauriault, 2014; Acker & Clement, 2019; Albury et al., 2017) (I examine the concept of data cultures in the second part of this study). Surveillance studies have investigated the words that either empower or disempower us when we discuss privacy, rights and democracy in the context of, for example, security and the technology we adopt to mitigate perceived risks (Lyon, 2015), and legal studies have investigated the discourse of law (Solove, 2002, 2001, 2008; Cohen, 2013).

In addition, the counter-narratives that feed into the development of alternative counter data cultures are crucial. In 2014, when Tranberg and I began to discuss and research for our book on data ethics as “a new competitive advantage” this was in fact one of the things we wanted to do: to trace and present an alternative narrative to a then-dominant discourse that values such as privacy were outdated and an obstacle to innovation (Hasselbalch, 2013, 2014). Back then, many were not convinced that data ethics was a term that would appeal to companies or they would tell us that we simply didn’t understand innovation. Nevertheless, by the end of the 2010s the “competitive advantage”-discourse was in fact overturning other “big data discourses” in public debate and in policy. As described in Chapter 1, the first data ethics group established by the Danish government was, for example, created with the specific objective to turn data ethics into a competitive advantage for Denmark. In addition, with a profound impact on late-2010s public discourse, Shoshana Zuboff deconstructed the dominant meta-narrative of the Big Data Society (2014, 2016, 2019) as one that was defined by powerful industries, and presented what she referred to as a counter “synthetic-declaration” that: “…value people, and reflect democratic principles.” (Zuboff, 2014).
Data ethics ‘critical cultural moments’ are possible when human memory and intuition are privileged and provided time and space to tinker.
ABOUT PART II

In this second part of the study on power and AI, I examine the history and special power characteristics of BDSTIs with AI capabilities, namely AISTIs. AI has progressed from the rule-based expert systems encoded with the knowledge of human experts and applied in primarily human and physical environments of the 1980s to systems evolving and learning from big data in digital environments with increasingly autonomous decision-making agency and capabilities in the 2010s. However, this evolution of AI does not imply, as I will illustrate in this part of the study, that also human involvement has been progressively excluded in the very data design; on the contrary, the degree of an AI system’s autonomy is prescribed by the human involvement in data processing from problem definition, collection of data, and data cleaning to training of the machine learning algorithm (Lehr & Ohm, 2017). Hence, the most widely used form of AI at the beginning of the 21st century, machine learning, can essentially be considered a data processing practice with a degree of autonomy that may be influenced by different forms of human involvement in the technical design, but also in its organisation and application in society.

In this part of the study, I also examine the ethical implications of AI (and thus the ethical implications of their evolution into AISTIs in the late 2010s) and the ethical theories that address these specific implications, narrowing them down to concerns with the relations between the distributed moral agency between human agency and the autonomous decision-making of AI, which is increasingly shaping our ethical experiences. I here consider an overarching theme within the AI applied ethics research field concerning different levels of human involvement in AI systems’ design, adoption and consolidation in society. These concerns, I argue, can also be traced back to ethical concerns regarding AI’s threat to humanity and human control or the potential of autonomous AI to surpass human deficits. However, I do not adhere to either one or the other poles of this discussion. Instead, I propose a middle way. Accepting that AI can be a moral agent, does not inevitably mean that we have to accept it is also an ethically responsible agent. That is, the question should not be whether machines should or can have human-level ethical agency and responsibility; rather, we need to focus on the ways
in which we can ensure humans continue to be involved in a meaningful and responsible manner. For instance, we can in very concrete ways create new standards and laws for AI and robotics to ensure such human involvement and empowerment in AISTIs (Pasquale, 2015, 2018, 2020). In addition, however, I propose that we also consider the very cultural systems in which AI gains its meaning as an either uncontrollable autonomous moral agent or as a human data design in which human involvement is a key property.

For this reason, I argue that a core ethical concern of a ‘data ethics of power’ is with AISTIs and BDSTIs’ constitution as cultural normative systems with a type of social ordering, in which interests of dominant actors in society have the primary advantage. I explore specifically the cultural systems that shape the practices of developers, scientists, lawmakers and users of AISTIs. Here, I more closely examine the interests in the data of AI as expressions of different data cultures with reference to the cultural systems of meaning-making that shaped the AI technological momentum of the late 2010s. I argue that a core ethical concern with AISTIs of a ‘data ethics of power’ is when human critical negotiation and ethical agency (the ‘critical cultural moments’) are immobilised in AI moral agency.

A key concern is with technological change as a field of power negotiation between the interests of different technological cultures, their compromises, or the dominance of one cultural system over another. With my point of departure in Hughes’ description of the “technological momentum” (1983, 1987), I thus propose adopting a helicopter analytical view on interests as a set of complex factors that come together in shared cultural knowledge frameworks and worldviews that cut across different stakeholder groups and communities.
CHAPTER 4

Artificial Intelligence
Sociotechnical Infrastructures (AISTIs)
AI IS EVERYWHERE. AND it is nowhere. What do we actually mean when we talk about AI? Is it a sophisticated improvement of our outdated human software? Is it a possible sci-fi scenario where an out-of-human-control machine out-competes humankind? Or, is it a commercial trade secret? Words are very powerful and, as abstract as they might seem sometimes, they actually have real consequences. Real laws are implemented based on the particulars of language; real business decisions are made; and real people’s lives are affected by the specific use of words and the worlds they portray. Evidently, the way we talk about AI defines what we think we can do with it and ask from it.

The founder of the singularity movement Ray Kurzweil believes that AI is the next step in human evolution:

“Biology is a software process. Our bodies are made up of trillions of cells, each governed by this process. You and I are walking around with outdated software running in our bodies, which evolved in a very different era.” (Lunau, October 12 2013).

The late scientist Stephen Hawking considered the power of AI an uncontrollable autonomous force:

“The development of full artificial intelligence could spell the end of the human race (...)It would take off on its own, and re-design itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn’t compete, and would be superseded.” (Cellan-Jones, 2nd December 2014).

The co-founder of Google, Larry Page, on the other hand, sees AI as just another (Google) service:

“Artificial intelligence would be the ultimate version of Google. The ultimate search engine that would understand everything on the web. It would understand exactly what you wanted, and it would give you the right thing. We’re nowhere near doing that now. However, we can get incrementally closer to that, and that is basically what we work on.” (Marr, 2017).

Whatever we say AI wants to be or can do for us will shape what we actually do with it and what role it will play in society (Hasselbalch, 2018).

In the mid 2010s, the term AI gained traction in public discourse, and particularly in business and technology companies, which started rebranding their big data efforts as AI (Elish & Boyd, 2016). Concurrently in
global policymaking, AI became a new item on the agenda of nations and intergovernmental institutions with the dedicated development of policy and investment strategies. With no shared definition, the term first and foremost was used generically to describe a sociotechnical evolution of big data technological systems. Amplified computer power and the vast amount of data generated in society had empowered machine learning technologies to evolve and learn to recognize faces in pictures (pattern recognition in images, ‘facial recognition’), to recognize speech from audio (pattern recognition in audio, ‘voice recognition’), to drive a car autonomously (rendering objects in an environment and perform a risk assessment), and to understand individuals when micro-targeting services and information (‘profiling’ and ‘personalisation’). These were all practical applications of AI systems increasingly adopted by companies and states to not only solve simple problems, and analyse and streamline disparate data sets, but also to act in real time, sensing an immediate environment and supporting critical human decision-making processes.

In this chapter, I examine the history and special characteristics of BDSTIs with AI capabilities, what I also refer to as AISTIs, their ethical implications, and the ethical theories that address these implications. The core objective of this chapter is to narrow down the data ethics of power considerations specific to AISTIs.
**CAN A COMPUTER THINK?**

Humans creating intelligent machines or life out of inanimate or dead things has been a theme throughout human history from the Greek myth of Deucalion and his wife Pyrrha, who made beautiful people by throwing stones over their shoulders, to literary and filmic depictions of the living corpse Frankenstein, the string doll Pinocchio and Metropolis’ humanlike machine Maria. Yet, the very scientific conception of a computation process with intelligence was most famously theorized by the mathematician and computer scientist Alan Turing, who in 1950 developed a method for testing a machine’s ability to have intelligent behaviour indistinguishable from that of a human (Turing, 2004).

The term Artificial Intelligence was, however, first coined in 1956 by the mathematics professor John McCarthy at the Dartmouth Summer Research Project seminar. He wanted to shift the focus of attention of computer scientists and mathematicians in the field of computation processes from the mere automation of these to the ‘intelligence’ of computers (Moor, 2006). Could a computation process do more than just process information and actually think information and learn from it like a human?

In the early AI research field, AI was explored by discerning the key differences (and similarities) between the human brain, feedback systems and digital computers (Crevier, 1993). As evidence of the similarities with the human mind and potential superiority of AI, chess playing computer systems were for instance later developed creating and acting according to game strategies. The most famous example was IBM’s Deep Blue that became the first computer to beat a chess champion when it defeated Russian grandmaster Garry Kasparov.

But fifty years after the Dartmouth seminar when five of the original scientists of the first seminar reconvened with other key people in an evolving and increasingly interdisciplinary field of AI research to discuss the next fifty years of AI development, the ambitions of the early AI research were more disparate (Moor, 2006). While McCarthy was now less ascertained about the creation of human level AI, others imagined AI with feelings and affect and the scientist and founder of the Singularity movement Ray Kurtzweil was certain that a Turing test capable AI was not
far away. The social science and psychology scholar Sherry Turkle, on the other hand, was less interested in the future potential of the intelligence of machines and more concerned with the human implications. (Moor, 2006)

One could argue that Turkle here represented a general twofold humanistic concern with the endeavour of the strand of AI science that sought to replicate the processes of the human brain and create intelligent non-human agents. Firstly, the relations between humans and machines alters human societies and minds in profound ways (Turkle, 1997). Secondly, we may add that the foundational questions regarding a computers’ intelligence and the undertaking to develop computer intelligence, and even consciousness, have from the outset been intertwined with concerns regarding what it means to be human and our unique status as the centre of our environments. Is the human neural system just another information processing system, complex, but also as material as the data processing of a machine? (Wiener, 2013, Bynum, 2010) And thus, if this is the case, is it even possible to argue that the human data processing agent (“inforgs”, Floridi, 1999) has rights that other non-human agents (also “inforgs”, Ibid. 1999) in our information environment (“infosphere”, Ibid. 1999) do not have? As Steve Woolgar puts it:

“Attempts to determine the characteristics of machines are simultaneously claims about the characteristics of non-machines (…) In discussing and debating new technology, protagonists are reconstructing and redefining the concepts of man and machine and the similarity and difference between them.” (Woolgar, 1987, p. 324).

The term AI has gone through several societal and scientific stages with different aspirations to create human-level AI or just very advanced problem-solving capabilities of computers. In 1980, the philosopher John Searle famously illustrated this fundamental conflict of views on the capabilities of AI in his Chinese Room example. He imagines that he is locked in a room where he is to respond to Chinese characters slipped under the door by following a computer program on how to do this. He does not understand Chinese, but by doing just like a computer does, following the program for handling the Chinese symbols, he can respond and slip back correct Chinese characters under the door, which convinces the ones outside the room that there is a Chinese speaker in the room. This example, he argues, illustrates the inadequacies of the Turing method. A computer may indeed create a satisfactorily response if it is programmed to act according
to the rules for interaction, but this does not mean that it is capable of understanding. Searle himself would not leave the room with an understanding of what was communicated to him through the door or what he responded himself. He therefore concludes from this example that strong AI has “little to tell us about thinking, since it is not about machines but about programs, and no program by itself is sufficient for thinking.” (Searle, 1980, p. 417).

Searle’s argument illustrates a conflict in the original aspirations of AI research to create respectively machines that think and understand by themselves or machines that ‘just’ process information and solve problems for humans. It also represents the early outlines of different sets of discourses that later would form essential frameworks for the development of AI research and adoption in society. Elish & Boyd (2017) describes AI as a technology that has always been suspended between the real and the imaginary cultural perceptions one being that of the agent machine that acts out of human control:

“…Western perceptions of what AI is – what it can and cannot do, and what it might yet do – are informed by long-standing cultural imaginaries of machines that escape the control of their creators, and the promises and perils of automata and artificial life” (Elish & Boyd, 2017, p. 8).

In the 2010s, the idea that machines may one day evolve entirely autonomously out of human control was still thriving; for example, Stephen Hawking warned in 2014 that the development to full AI could end mankind (Cellan-Jones, December 2, 2014); moreover, the founder of the Singularity movement Ray Kurzweil predicted in 2016 that AI will update the outdated software of humanity to create an entirely superior intelligence (Cellan-Jones, 2016). These ideas can also be traced in more general public discourses on AI intelligence and potential agency of machines, and accordingly in imaginings about the imminent potentials or threats of AI. Concerns regarding AI agents that replace the human labour force or the artistry and creativity of new AI systems represent the imagining of an autonomous new agent in society.

As such, it may be argued, as Elish and Boyd (2017) do, that the “magic” of AI is only a mystification of a technology that becomes part of “hype” and “fear” cycles, which in the end disempower us in what we think we can do with AI. They therefore also argue that these cycles may only be countered
by developing a rich methodological framework for data analysis referring to the very design process of AI. One may also extend this argument to a data ethics approach to the development and adoption of AI in society.

However, to develop a methodological framework to do this, we need a conception of AI as a digital data process that can be designed and governed by humans. That is, narrowing our focus on AI as designed data systems and data processes makes AI more manageable than governing a rogue independent agent in society. AI’s gradual practical implementation in society has progressed from rule-based expert systems encoded with the knowledge of human experts, applied in primarily human and physical environments in the 1980s, to systems evolving and learning from big data in digital environments with increasingly autonomous decision-making agency and capabilities in the 2010s. It is also this latter practical application of AI as digital data processing that I use in this study.

4.1 Expert Systems

The history of the technological development of AI consists of social peaks and lows predominantly due to its different levels of practical commercial application and philosophically challenging ambitions. In his account of the history of the development of AI from the 1950s to the 1990s, the AI researcher and entrepreneur Daniel Crevier (1993) describes the endeavour to artificially construct intelligence as a thrive also to uncover the complex essence of human thought. This was not a modest ambition, and AI research in the 1950s and 1960s was first and foremost experimental performed inside research labs with different aspirations to imitate human decision-making and thought processes in math and computer processing. Consequently, in the mid 1970s the AI research field experienced its first ‘AI winter’, where these original grander ambitions lost traction in funding and investment communities due to a lack of practical application (Crevier, 1993).

However, in the 1970s, the development of logic-based programmed ‘expert systems’ also created a new space for an initial commercial adoption of AI. As a consequence, in the 1980s expert systems were created to support or even replace decision-making in professional settings, where information was collected from human experts and then coded as rules and procedures of the computer (Alpaydin, 2016).
The promises of these systems to reduce the costs of human resources were at first very grand. Crevier provides different examples from industries in the 1980s that replaced human experts with expert systems with the aim of reducing the cost of training people and moving field experts around to share their knowledge and, for example, troubleshoot problems. One example was the North American General Electric Company whose experienced engineer David Smith was the only person who could handle electric locomotive repair problems, and who would therefore be physically transported around to fix broken engines. In 1981, when Smith was considering retirement, the General Electric company managed to codify his expertise into an expert system named the Diesel Electric Locomotive Troubleshooting Aid (DELTA). It contained hundreds of rules for troubleshooting and help, representing the knowledge of Smith; in 1984, DELTA could diagnose 80% of the breakdowns and provide detailed instructions for performing repair on broken engines (Crevier, 1993, p. 198).

The expert systems of the early 1980s were promising in their prospects to reduce costs and distribute and sustain expertise within a company. However, many also soon proved to be less valuable, working only in limited settings and with results left wanting (Alpaydin, 2016). In DELTA’s case, users were supposed to take over the maintenance of the system after its initial development, but no one wanted to take on this responsibility and it was therefore never used (Gill, 1995, p. 66). Some of the problems with the early expert systems were caused by the development of technical environments, such as an expert system being misaligned with a company’s general computing environment (Gill, 1995, p. 64). However, other problems with the systems could be traced back to their inability to adapt to human environments, as was the case with DELTA; for example, concerns about the liability of developers and companies using the systems, problems solved by the systems not being considered critical by users, users’ resistance to externally developed systems, or the loss of key developers of the systems (Gill, 1995).

### 4.2 Machine Learning Decision-Making Systems

What happened in the years following the creation and application of the first expert systems is to a large extent also a story about the development of
an increasingly digitalised big data environment, which enabled what is also referred to as ‘machine learning’ systems. Machine learning was the most practical application of AI in the 2010s. With machine learning, an AI system’s knowledge essentially no longer has to be provided by human experts as the system will learn and evolve with data; accordingly, the system gains its autonomy and agency. Here, it is no longer the human expert that capacitates AI but rather digitalised data sets; a machine learning system learns with and further evolves on the basis of automated data processing.

Lehr and Ohm (2017) describe the data analytical capabilities of a machine learning process as “...an automated process of discovering correlations (sometimes alternatively referred to as relationships or patterns) between variables in a data set, often to make predictions or estimates of some outcome.” (Lehr and Ohm, 2017, p. 671). One example is Apple’s Siri, which analyses verbal questions and orders either directly on the device or by searching the internet. What makes Siri intelligent is that the program evolves and learns from data one provides through the questions one asks. In this way, the program assimilates itself by creating a data profile of the user and eventually becoming more of a personal assistant.

In the late 2010s, machine learning systems for analysing and acting on data in real time were increasingly embedded in data systems in all societal sectors and spheres, from healthcare to social networking. Professor of computer engineering Ethem Alpaydín (2016) credits this revolution of machine learning systems to the creation of the digital environment. In the 1980s, the invention of the microprocessor initiated the massive development of personal computers, and thus, computers and later personal devices were distributed widely in populations. The digitalisation processes of the 1990s going into the 2000s further enabled a pervasive massive collection of big data. Now all information from colours in a photo to tones in an audio recording could be transformed into a set of numbers and processed by computers. (Alpaydín, 2016)

These technological developments paved the way for the fast-paced advancement of machine learning and a growing portfolio of internet-connected things, further capacitating increasingly autonomous behaviour and analytical agency in AI systems that learned and evolved via big data. To illustrate this, the Cognitoy Dinosaur was a toy that used one of the most powerful machine learning models in the world, the Jeopardy-winning IBM Watson computer, to assess a child’s interaction with it. The toy was not
programmed with predetermined responses, but rather learned from a child’s questions and responses and tailored its own responses to them. A child would say “my favourite colour is red” and the dinosaur would respond “okay, I will try to remember that”, while storing this information for future more personalised play.\(^\text{10}\)

Notably, while machine learning cut out the human expert in some aspects, it did not entirely exclude human involvement in the very data design of machine learning systems. On the contrary, the degree of an AI system’s autonomy is prescribed by human involvement in data processing, from the problem definition, collection of data, and data cleaning to training of the machine learning algorithm (Lehr & Ohm, 2017).

Lehr and Ohm (2017) refer to this human involvement in machine learning processes as “playing with data”. They argue that legal scholars have been too focused on the autonomy of machine learning when primarily concerning themselves with the “running model” of the systems (the way they are adopted and used), while neglecting the data processing activities that shape a machine learning system. Machine learning algorithms, as they state, are not magical black boxes with mysterious inner workings. In fact, they are the “complicated outputs of intense human labor — labor from data scientists, statisticians, analysts, and computer programmers.” (Lehr & Ohm, 2017, p. 717). In this perspective, we may consider the most widely used form of AI in the beginning of the 21st Century, machine learning, as essentially a data processing practice with a degree of autonomy that may be influenced with different forms of human involvement, not only in the technical design but also in the organisation and application in society.

It is also this conception of AI that I use in the present study. Rather than a scientific or technical term, I here consider the concept’s revival and application in a specific moment in history and do not ponder further on the potential ‘intelligence’ of AI (technologically or philosophically). I use the term in this way to address the more generic use of the term in European policy discourse in the late 2010s concerned with the technological and social evolution of big data systems in society. I also want to argue that while the term AI indeed was the term most commonly used in this period in policymaking, artificial human-level intelligence, as such, was in fact not the

\(^{10}\) See CogniToys’ promotional video about the dinosaur: https://www.youtube.com/watch?v=HPw0jIxr2kk
emphasis.

One of the first tasks that the HLEG on AI set out to accomplish was to create a definition of AI. This technical definition that was later published as an official deliverable of the group emphasised the data processes and human involvement in AI:

“Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal.“ (HLEG C, 2019, p. 6).

The policy and investment strategies of the European Commission that followed were also created to specifically secure the data resources of AI (e.g., as expressed in the EU’s data strategy published February 2020, European Commission H, 2020) or to ensure the development and access to big data AI enhanced tools (products and services) to support European private and public sectors (e.g., in the EU’s AI white paper published together with the data strategy in February 2020, European Commission I, 2020).

4.3 Artificial Intelligence in Society

Now, we have a technical definition of AI as a complex data processing system with a level of autonomy shaped by human involvement. AI enhances the analytical and sensory technical capabilities of big data systems distributed in networks via a myriad of digitally connected devices and things. However, these new technical capabilities are also a societal evolution as they are embedded in the very sociotechnical infrastructure of societies, thereby transforming sectors and the private and public sphere of citizens in profound ways. Crucially, in the 2010s, the private and public sectors saw a fast-paced adoption of AI systems. AI applications were created for education, the environment, energy, healthcare, policy, financial IT, smart cities, mobility, and sustainability among other areas (Allam & Dhunny, 2019). If not implemented, strategies for their adoption in different sectors were developed with an emphasis on different degrees of human involvement. The following are some of the examples from different societal sectors of the 2010s:
- **The public sector**
  In different European countries, strategies were created to for instance integrate AI in public institutions to develop personalised assistance, chatbots and conversational platforms, as well as for socially scoring families and tracing vulnerable children. Moreover, the public sector saw the use of applications for automating civil servants’ tasks, predictive policing, and fraud detection (Spielkamp, 2019).

- **The financial sector**
  In the financial sector, finance was transforming into “cyborg finance” in which humans and machines “share power” (Lin, 2014); for example, 80% of transactions on the Forex market (where the world’s currencies are traded) were performed by robots (Bigiotti & Navarra, 2019). Finance ‘Robo advisors’ provided financial advice for investment management (Lieber, April 11 2014) and financial institutions used AI applications for market analysis and to assess credit quality and price loan contracts (Financial Stability Board, 2017); moreover, AI systems were used to assess and credit-score consumers (Pasquale, 2013).

- **Social networking**
  The most common social networking services, provided by platforms such as Facebook and Google, were using AI systems to provide personalised ranking and recommendations, analyse the words and phrases of search queries and decide what other words have the same meaning, provide facial recognition to track users in photos, understand and respond to conversations, or detect misinformation and illegal and harmful content.¹¹

- **Smart cities**
  Cities were transformed with internet-connected things with sensors that collected data in real time. AI was integrated in city management and engineering and construction to analyse distribute in real time and centralise data from different urban components. (Allam & Dhunny, 2019)

- **Health care**
  In health care, AI was used to support decision-making in the probability

---

¹¹ Facebook and Google's AI sections: [https://ai.facebook.com/](https://ai.facebook.com/), [https://ai.google](https://ai.google).
and estimation of diseases, personalised medicine, illness monitoring and treatment planning, critical care, diagnosis, treatment decisions and triage. (WHO, 2018)
From a 1950s’ scientific endeavour and sci-fi curiosity, AI had by the 2010s transformed into a young sociotechnical system with rapid societal adoption. As described, the new technical AI capabilities of big data systems were also a societal evolution. The systems of the infrastructures of private and public sectors were gradually transforming: they were becoming less physical and human-controlled, and increasingly based on digital big data and enhanced through AI. As follows, decision-making processes were in these sectors increasingly also informed by and even also replaced by big data AI systems of prediction and different types of risk or potential analysis.

What did this mean in practice? Let us examine a few examples: Recommendation and personalisation systems profile and analyse our personal data and decide for people what they see and read and who they engage with online. Systems for autonomous driving scan the street in front of the driver, evaluate the risks and worth of the different objects in the way, and decide who or what the car hits in case of an unavoidable accident. Judicial risk assessment systems look for patterns in backgrounds of defendants to inform a judge about who is most likely to commit a crime in the future. Triage systems process the medical and demographic history of patients to decide who gets the kidney. All of these processes of partly or primarily autonomous decision-making of AI systems comprise ethical dilemmas that are increasingly extended into AI systems. By way of illustration, as citizens in a democratic society, what exact level of choice and insight into the information of the internet should we have? Or, on a very basic level, who should the car hit? The young person with a criminal background or the elderly person who never committed a crime?

As AI systems were envisioned, adopted and embedded in societal infrastructures in the 2010s, their ethical implications were also materialising in the shape of moral decisions and choices intertwined with the complex data processing of AI systems. Hence, concurrently with a renewed public focus on AI, ‘AI ethics’ – or ‘Algorithm ethics’ – emerged as a research field concerned with the ethical implications of AI systems.

Although different terms are used, I here use the general term ‘AI ethics’ to describe an emerging research field that addresses areas of concern in regards to the ethical implications of the practical application of AI in
society. Thus, David Leslie (2019) from the Alan Turing Institute describes ‘AI ethics’ as follows:

"In order to manage these impacts responsibly and to direct the development of AI systems toward optimal public benefit, you will have to make considerations of AI ethics and safety a first priority.” (Leslie, 2019, p. 3).

In the following section, I examine how the ethical concerns I voiced in the previous section regarding different levels of human involvement in AI systems’ design, adoption and consolidation in society, can also be traced as an overarching general theme within the ‘AI ethics’ research field. Furthermore, these concerns can be associated with the aforementioned imagined scenarios regarding AI’s threat to humanity and human control or the potential of autonomous AI to surpass human deficits. As I will illustrate, due to different aspirations and conceptions of AI, the practical application of AI ethics also deals with very different levels of human involvement in the design and governance of AI. In my view, the most valuable applied AI ethics approach is the one that prioritises the highest level of human involvement in AI development.

The ethical implications of AI systems’ role in human and societal decision-making processes are also a general theme in the AI ethics debate. It spans, as I will show in the following sections, from discussions regarding the role of machine agency in the moral world of humans as either a positive or destructive transformative force, to applied AI ethics’ methodologies and frameworks. The latter in particular, can be examined from the view of embedding different degrees of human agency and involvement in the very design and organisation of AI decision-making processes.

While I consider a perspective on the very design of AI a constructive and highly relevant contribution, I also propose that it is only one of the applied ethics components of an ethical governance framework for AI systems. I thus propose that a data ethics of powers’ contribution to the AI ethics debate is a reflection on human power in AI development; that is, the different levels of human involvement we assume and request in the development and adoption of AI at a more general level. Accordingly, to narrow the discussion on AI ethics to a ‘data ethics of power’, in the following section I primarily concentrate on the levels of autonomy and human involvement in AI systems’ data processing. Here, I address not only the developer side of data design but also the social and cultural adoption and consolidation of AI data.
4.4 From Human-Dependent Systems of Expertise to Autonomous Systems of Decision-Making

AI systems have, in their short history of societal adoption, been used to support or replace human decision-making processes with various levels of autonomy. The agency and autonomous behaviour of AI systems were in their practical application not an objective per se, but a fundamental feature of the system’s ability to adapt to real-life decision-making processes.

In the 1970 and 80s, expert systems were created to aid humans in decision-making, such as when troubleshooting and guiding decisions on how to repair machines or diagnose infectious diseases (Crevier, 1993). They were composed of a “knowledge base”, which consisted of a range of facts and “if then rules” based on the knowledge of human domain experts, and an “inference engine” using logical inference rules to deduct new knowledge (Alpaydin, 2016, p. 50). Previously, I described these early expert systems’ inability to adapt to human environments as a key reason why they did not succeed in societal adoption. This also included the way in which they represented the real world. The logical rules of the systems were simply too rigid to represent it, they could not represent the nuances and graduation of life. Alpaydin uses the example of age; one is not just ‘old’, but we are growing old gradually, and this process cannot be captured by a figure (Alpaydin, 2016, p. 51). The early expert systems’ evidence (via pre-programmed knowledge and logical rules) was evidently very limited and often also faulty in terms of the representation of the nuances of a real environment. They were thus failing in presenting valuable decisions and their practical application was for this reason very limited.

Contemporary AI systems are extensions and improvements of these

12 Here, we could use other examples from the more present world of social media content moderation, such as the meaning of concepts like “indecency” and “harmful”. Needless to say, time and again the most publicly controversial cases of content blocking and take-down on social media portals in the 2010s stemmed from different interpretations of what constitutes “harmful content”, and critically the social media portals’ automated content moderation systems’ decision-making power and thus enforcement of specific politics and values in the public online sphere. See also the European Commission’s original 1996 distinction between “illegal” and “harmful” content online and accordingly the different legal responsibilities associated with content as such https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:1996:0487:FIN:en:PDF
original expert systems in decision-making. However, increasingly they can reason, make decisions and learn by themselves via complex multi-layered data processing and sensors, which make them capable of perceiving complex environments (HLEG, C, 2019, p. 3). They are therefore also better (but of course never perfect) at analysing the nuances of real-world settings (Alpaydin, 2016, p. 52). Machine learning, which uses algorithms based on the concept of neural networks, such as in deep learning, dynamically uses input data from sensors, which is then processed progressively in a way where each layer of analysis takes its input from the previous to produce a decision (Alpaydin, 2016, p. 85).

4.5 Moral Machines

Now, it is one thing to build evidence for the troubleshooting of a faulty engine and make a decision regarding its repair—it is another thing altogether to build evidence with the nuances required to make complex moral decisions that affect human lives. Increasingly, AI systems are implemented in settings that involve ethical reflection and moral decision-making per se or that are transformed by the systems in ways that produce new ethical implications.

Most famous is the autonomous vehicle ethical dilemma presented by Awad et al. (2018) in their “Moral Machine” experiment that exemplified the moral decisions involved in connection with driving a car in an accident involving pedestrians. By way of illustration, if the only choice is between causing the car to crash with the driver, hitting two elderly people or hitting one young person, what would be the morally right decision to make? Awad et al. developed an online serious game with ethical dilemmas of different scenarios such as this to examine people’s moral choices to create “…a global conversation to express our preferences to the companies that will design moral algorithms, and to the policymakers that will regulate them.” (Awad et al, 2018, p. 63). They focused on the “running model” (Lehr & Ohm, 2017) of the AI system’s decision-making process; that is, the moment it is deployed and makes a decision entirely on its own without human involvement. In this way they also imagined an everyday life in the future where machines will replace human decision-making and act autonomously:

“Never in the history of humanity have we allowed a machine to autonomously decide who should live and who should die, in a fraction of a
second, without real-time supervision. We are going to cross that bridge any
time now, and it will not happen in a distant theatre of military operations;
it will happen in that most mundane aspect of our lives, everyday

The “Moral Machine” experiment reimagines the famous trolley
problem, where different scenarios and ethical dilemmas are tested in
practice. It has become the most used example of AI ethics; however, I do
not consider it the best one in the context of understanding the practical
implementation of a human (-centric) approach to AI. In fact, the very
programmed choice of the machine is not the ethical dilemma we want to
consider first. I want to start before that; to think about the ethics of a
machine that does not make autonomous decisions without us. What we
want to ask is this: how do we want the machine to help us make the decisions
we as humans want to make? How does the machine complement a human
environment? In this way we could think of a future alternative to the one
envisioned in the “Moral Machine” experiment in which no critical decisions
can be made without human involvement. In European law (General Data
Protection Regulation Regulation (EU) 2016/679) there are, for example,
provisions that prohibit decisions regarding individuals based solely on
automated data processing without human involvement that significantly
affects the individual. (Regulation (EU) 2016/679, Article 22).

4.6 Ethical Implications of AI Decision-Making

In assessing and mapping the ethics debate on algorithms, Mittelstadt et al.
(2016) identify different types of ethical concerns connected with the way in
which algorithms process and make correlations in data (make evidence out
of data) to reach decisions. The ethical implications, such as discriminatory
decisions, can be immediately “visible” outcomes, actions that can be
discerned as “unfair” in the moment of observation. However, ethical
implications may also be societally transformative in ways that are not
observably harmful in the moment of implementation. (Ibid., p. 5). Here, the
challenges to autonomy are an overarching concern that considers most
predominantly personalisation algorithms and the construction of “new
choice architectures” which may nudge our behaviour and control our
decisions to different degrees (Mittelstadt et al., 2016, p. 9). Concerns about
the challenges to the way in which we conceive of and deal with informational privacy brought about by big data collection and processing in the form of profiling algorithms, for example, is another. Lastly, they consider a horizontal concern regarding the traceability of algorithms (Mittelstadt et al., 2016, p.12). The very design of the complex data processing of algorithms’ evidence is often difficult to trace and therefore generally complicates the identification of responsibility for ethical implications of algorithms. However, the level of human oversight of increasingly complex systems may also be complicated by other factors. A lack of education or other human level capacities (such as one’s level of awareness and ethical reflection, and cultural and social factors) may equally make it more difficult for humans to identify and/or correct design, which leads to ethical implications. Lack of human oversight also means that control and agency are gradually moved into the system.

4.7 Machine Ethics

The AI conversion of human–machine relations raises ethical concerns in regards to human agency and involvement in systems of increasingly distributed moral decision-making. However, a future characterised by autonomous AI moral decision-making is not only conceived of as a potential future scenario in the ‘AI ethics’ research field. ‘Machine ethics’ (Anderson & Anderson, 2011) focuses on the ethical behaviour of autonomous AI agents. A foundational prediction here is that in the future, human involvement will be minimal, and therefore machines must have ethical and moral capabilities. As follows, we need to develop theories and methodologies to train machines to act ethically:

“Theoretically, machine ethics is concerned with giving machines ethical principles or a procedure for discovering a way to resolve the ethical dilemmas they might encounter, enabling them to function in an ethically responsible manner through their own ethical decision making.” (Anderson & Anderson, 2011, p.1).

Within the machine ethics research field, AI agents with ethical capabilities might even improve human moral decision-making. Intelligent machines are considered morally superior to humans (Dietrich, 2011, Anderson, 2011). They can help us create universal ethical principles by surpassing the relativism of a human moral decision-making driven by a self-serving interest (Anderson, 2011). Following this, Seville & Field (2011)
envision an “ethical decision agent” that can help people make ethical decisions by pointing to consequences of their decisions or a virtual reality to create “ethical experiences”. This agent, they argue, would be more impartial and add consistency to moral decision-making.

The literature with perspectives on ‘machine ethics’ provides valuable insights in regard to the design and implementation of technically mediated components of moral decision-making, which is increasingly distributed in external technological systems. The recognition of a new type of technological agent that actively participates in moral decision-making processes, and consequently also in shaping our ethical experiences, is of particular importance here. However, as I argue in the following section, AI agents do not take on moral responsibility as more impartial moral agents imposing universal moral norms. Moral responsibility, I maintain, will always be intrinsically part of human involvement in the design, development and implementation of AI. This also implies that responsibility for any moral action and ethical implication of an AI system can and should only be that of the humans involved (Bryson, 2018). As such, the very argument for building artificially moral agents with superior moral skills to correct the errors of human moral reasoning is also challenged (Wynsberghe & Robbins, 2019).

4.8 AI Moral Agency and Human Ethical Responsibility

In 2017, the European Parliament adopted a resolution with recommendations on Civil Law Rules on Robotics. In this resolution, the question of legal liability for harmful actions and implications of autonomous systems was raised. “Consider”, the resolution therefore states, a “new legal status” for autonomous robots, possibly an “electronic personality” (European Parliament, 16th February 2017). I believe that even considering a responsibility as such of AI agents has dire ethical implications as it implies that we also accept autonomous ethical agency. I insist that we do not have to do this, because human involvement and agency is, although at times difficult to discern, always present in AI. It is present in the how of the data of AI, as I have illustrated in the previous sections; in the laws that frame AI; in the technological cultures of its design; in the way in which we handle and adopt AI in society; and crucially human agency is present in
how ‘AI autonomy’ is socially perceived, accepted or rejected. Thus, while AI does indeed have technical decision-making capabilities and may be imagined in the context of autonomous machine agency, human involvement always plays a role. Accordingly, what has to be done is to enhance and support this ‘human factor’ in the design, development, adoption as well as legal frameworks for AI.

In the ‘AI ethics’ debate, two opposite poles represent the threat to human moral (ethical) agency and control, or the potential of powerful and humanly superior machines for human moral (ethical) improvement.\footnote{See my distinction between ‘moral agency’ and ‘ethical agency’ in Chapter 6 and the Terminology section for further elaboration.} However, there is a middle way. By accepting that AI systems are moral agents, I maintain that we do not simultaneously acknowledge that they are also ethical agents; that is, \textit{ethically responsible} agents. All that is recognised is that humans are not the only agents that shape the moral architectures of the environments in which we live (Adams, 2008). Said in other words, AI does not cause a “responsibility gap” where blame for ethical implications has no object (Tigard, 2020). There is always an ethically responsible human dimension to AI. By way of example, we can think of the ethical implications of the application of an autonomous vehicle that kills a pedestrian in terms of the complexity of the human conditions that led to the accident; as the result of a network of human processes - from design choices to implementation, and even the adequacy of the laws that frame the development and the use of the autonomous vehicle as well as the rules and the shape of the streets on which it drives. This does not mean that AI systems escape legal accountability; all it means is that only humans can be ethically responsible. As the philosopher Daniel W. Tigard points out, moral responsibility and legal accountability are two different things:

“By looking beyond accountability, we see that some who are harmed by technology will want to better understand the reasons for a machine’s behavior or the underlying values that it seems to have been programmed or learned to promote.” (Tigard, 2020, p.14).

Now, if we consider AI decision-making systems as components of socio-technical architectures of distributed moral agencies where humans and nonhuman agents are intertwined in shaping moral experiences, questions regarding the moral status of nonhuman agents become more practical than
existential in nature. In this case, we do not even need to ask if machines should or can have human-level ethical agency and if they can be ethically responsible, but rather we should attempt to find a way to ensure that humans continue to be involved in a meaningful and, crucially, responsible manner.

Here we may use, as I have done in the third article of this study on AI and data interests, Bruno Latour’s (1992) description of the “moral agency” of technological artefacts as delegated nonhuman actors that enforce human laws (which include “…values, duties, and ethics.” (Latour, 1992, p. 157)). He argues that technological artefacts are “strongly social and highly moral” (Latour, 1992, p. 152) working by prescription (of laws and orders that are “…inscribed or encoded in the machine”, (Latour, 1992, p. 177)). By way of Latour’s illustration, a technological artefact such as a seat belt can lock our bodies in positions we do not wish to be in. It is designed to do exactly this, does indeed enforce the laws of car safety, and will certainly let us know with an insistent beeping sound if we are not ascribing to these laws.

However, technologies are not just passive expressions of moral intentions; they are what Latour refers to as “technical mediators” (Latour, 2002, p. 252). Moral intentions and actions are actively translated in the technical design intertwined with various possibilities that are in constant negotiation with use, laws, culture and the society in which we act. We can here use the example of ‘word embedding’ machine learning methods, which are used for the language processing of online search engines. Examining the most commonly used models, Bolukbasi et al. (2016) found that word clusters were created in which words such as architect, philosopher, financier and similar titles were grouped together semantically as “extreme he” words, whereas words such as receptionist, housekeeper and nanny were grouped together as “extreme she” words (Bolukbasi et al, 2016, p. 2). Hence, they conclude that the blind application of this model could contribute to discrimination in society. The work of a developer of this type of machine learning model clearly constitutes a moral action that entails the delegation of moral values and choices to a technological moral agent that then shapes the moral factors of a social environment. As Bowker and Star (2000) argues classifications of information are never neutral. There is always a “moral” dimension (Bowker & Star, 2000, p. 5) to the work of a developer of an information system, such as the word embedding model of a search engine. Thus, this human actor has a crucial role as the actor who
inscribes the programming language of the technology’s delegated moral agency (Latour, 1992). However, the translation of this type of technological moral agency into moral action and/or implications is not a straightforward process. If we trace the moral agency that in this example results in societal ‘discrimination’, several active actions are not solely those of the human developer of the machine learning model neither are they solely those of the ‘machine’. The machine learning model (nonhuman actor) actively amplifies existing human bias (human actor) in its training data; that is, it is learning from and evolving (creating gender-biased word clusters) from Google news articles. Nevertheless, it is also ‘blindly’ developed, accepted and enacted in society as an objective representation of information by human actors (the user and the developer).

We need to examine the relations between the distributed moral agency of active human and nonhuman actors. Social and ethical implications are in this perspective not just the result of a human intention, neither are they of nonhuman actors. Rather, they are consequences of a network of actions and competences distributed between these different agents. What does this model of distributed moral agency then mean in terms of applied ‘AI ethics’? How do we consider the ethical implications and actively apply ethical considerations in the development of AI systems? Firstly, it means that we cannot just design a moral norm into a technology and thus produce a "Moral Machine"; we need to address the way in which ethical implications evolve in environments of distributed moral agencies between human and nonhuman actors.

When describing how to design “artificial morality” in artificial agents, Allen et al. (2005) offer two approaches. One is the top-down approach, in which the machine is designed to act according to specific moral principles; that is, moral theories may be used as the programmed rules for the selection of ethically appropriate actions (Allen et al., 2005, p. 149). We return again to the “Moral Machine” experiment, which has the objective of creating a global foundation for designing an autonomous machine that may take moral action by itself based on the moral norms inscribed by humans. This model deals with the extension of the moral intentions of a human global society into external technological systems. However, the model does not take into account the type of dynamic distributed moral agency that I have just described. This is where we may use Allen et al.’s (2005) second approach, where we do not impose a specific moral theory.
but aim to provide environments (with for example meaningful human involvement and agency) for AI agents in which appropriate behaviour is selected and rewarded (Allen et al., 2005, p. 151). In this way, the machine acts ethically by dynamically evolving in ethical environments. This description of a critical applied ethics approach to AI leads to the following discussion of the contexts of power relations and interests that AI technologies evolve in.

4.9 Interests and Power Relations

The AI ethics research field comprises a recent concern with the ethical implications of increasingly autonomous data systems and algorithms. However, the discussion merges with previous debates regarding the neutrality of computer technologies. The conceptualisation of the entrenched values of a computer technology design was originally formulated by Batya Friedmann et al. in the 1990s, and has since then been further explored in the value sensitive design (VSD) framework (Friedman & Nissenbaum, 1995, 1996, 1997; Friedman, 1996; Friedman et al., 2006, Flanagan et al., 2008; Umbrello, 2019; Umbrello & Yampolskiy, 2020; Umbrello, 2020). In this perspective, a computer technology is never neutral, but rather embodies moral values and norms in its very design (Flanagan et al., 2008).

In VSD the embedded values of a technology are addressed as ethical dilemmas or moral problems to solve in the very design and practical application of computer technologies. The ethical implications of a computer technology can therefore be analysed by examining the technical design, which can be designed in ‘ethical’ or ‘ethically problematic’ ways. Friedman and Nissenbaum (1996), for instance, illustrated different types of bias embedded in existing computer systems used for tasks such as flight reservations or the assignment of medical graduates to their first employment, and presented a framework for addressing this in the design of computer systems.

The VSD approach has been similarly employed in studies of the values embedded in AI system design extending the analysis to the entire life cycle of AI (Friedman & Hendry, 2019; Umbrello & Yampolskiy, 2020). The data systems and mathematically designed algorithms of AI are not impartial or objective representations of the world. Consequently, they also trigger actions
and societal effects (decisions and suggestions) that are not “ethically neutral” (Mittelstadt et al., 2016, p. 4).

One set of research of the late 2010s specifically addressed the ‘non-neutrality’ of AI systems with specific reference to the power dynamics of the adoption and implementation of AI and big data systems in the public and private sectors of the period. Several concrete examples and case studies of social and ethical implications of AI systems were here used to highlight the social power relations and interests at play in the development and societal adoption of AI.

Cathy O’Neil (2016), for example, described the darker side of the big data systems, or what she referred to as “Weapons of Math Destruction” (WMDs), which in the 2010s were implemented in the U.S. educational and public employment systems for credit score and insurance assessments. Her primary concern was that the big data systems are deployed without questioning and assessment of their social implications by private and state actors as neutral and objective systems to replace human decision-making and assessment. She illustrated the at-times devastating consequences for citizens. For example, teachers were let go based on rigid machine-based performance assessments that do not take into account social contexts and human factors (O’Neil, 2016, p.5), and furthermore, people from less desirable demographics received lower credit scores based on their computer’s location (O’Neil, 2016, p. 144).

Frank Pasquale (2015) worried equally about the use of automated processes to assess risks and allocate opportunities. These are controlled by private companies, which he argued are also the most profitable and essential parts of the information economy (Pasquale, 2015, p. 216). Complex algorithms are developed and deployed to sustain a “Black Box Society” where the data processes of algorithms are protected intentionally as trade secrets to sustain information monopolies of powerful industries. These industry interests also authorise computers to make decisions without human intervention.

Another famous case study of the power relations at play in the adoption of an AI system was the “Machine Bias” study published by the news site Propublica (Angwin et al., 2016). Here, the investigative journalist Julia Angwin together with a team of journalists and data scientists examined the private company Northpoint’s COMPAS algorithm, which was used to perform risk assessments of defendants in the U.S. judicial system and assess the likelihood of recidivism after release. They found a bias against black
defendants in the algorithm that had the tendency to designate black defendants as possible reoffenders twice as often as it did white defendants. Moreover, it grouped white defendants rather than black defendants as a low risk more often.

Safiya Umoja Noble (2018) described the power of the search architectures of Google’s algorithms as oppressive (“algorithms of oppression”). They not only reinforce bias in society when they, as she illustrated in her investigations, replicate prejudices against for example black girls in sexualised and discriminating search results. They enact bias when presented and received as natural objective categories of reality.

The ability to examine the potential discriminatory agency of an algorithm is essential in a society in which ethical implications arise from the distributed moral agency among human agents (developers and, e.g., judges that use a risk assessment tool) and nonhuman agents (the tool’s data design, e.g., Google’s search algorithms). We can even offer critical applied ethics methodologies for correcting the discriminatory agency within the very data processing of the machine learning model of the tool. Lehr and Ohm (2017), for instance, have presented different ways of intervening in the “playing with data” stages of a discriminatory machine learning model; for example, by examining the way in which it facilitates the translation of disparities in training data into prediction disparities, generating less accurate predictive rules for minority groups than for others (Lehr & Ohm, 2017, p. 704). However, this can only be done if we have the option to trace the actual data processes of the algorithm. Mittelstadt et al. (2016) argue that the traceability of algorithms is complicated by the increasingly complex data processing of algorithms and humans’ capacities (such as awareness and education) to identify and/or correct the data design of an algorithm.

However, another angle exists. Angwin and partners (2016) had to provide their evidence for the “Machine Bias” study without accessing the ”playing with data” (Lehr & Ohm, 2017) of the COMPASS software as its algorithm was protected as proprietary of the private company behind Northpoint. Their critique of the system was complicated by the commercial interests of the private company behind. Thus, they could only study the “running model” (Lehr & Ohm, 2017) by comparing different data sets with public records requests (Larson et al., 2016). This very inaccessibility and ‘autonomy’ of a socio-technical tool, enforced by humans with the interests of a company, are ethical implications by themselves. It is here not just a
question of technical complexity and human education and capacity, but also the result of societal power relations expressed in legal protections or freedoms. The lack of traceability can therefore not be reduced to a clash between an autonomous, technically complex moral agent and the individual human’s capacities. As Pasquale argues, and the “Machine Bias” case study illustrates, the general social power relations and interests at play also complicate audits and interventions by preventing access to the very “playing with data” of the “running model” (Lehr & Ohm, 2017).

4.10 Data Ethical Governance of AI

In summary, ‘AI ethics’ is first and foremost an applied approach to the development of an AI system’s design; however, without addressing the distributed moral agency in the context of the general power dynamics of a society, it is to a certain extent futile. As argued in part I and further expanded on in this chapter with reference to the special characteristics of AI, to guide the direction of a sociotechnical system, we must address its complexity with a multi-actor, reflexive, open-ended (Hoffman et al., 2017) and agile approach (Winfield & Jirotka, 2018).

In this perspective, which type of critical applied data ethics (‘data ethical governance’) for AI may we suggest to for example policymakers? I have in this chapter addressed the autonomy of AI systems as an ethical problem to solve. Subsequently, I have examined the level of human involvement in AI development and adoption as one key component to consider in an ‘AI ethics’ applied framework. Here, I have also focused specifically on the data processing of AI. I have examined the level of autonomy and thus human involvement in AI as something that can be addressed in the micro contexts of developing AI; that is, in the very “playing with data” (Lehr & Ohm, 2017), but crucially, I have also examined the autonomy of AI systems in the context of human societal power dynamics; that is, the level of AI autonomy as something that is shaped by different interests in society. At last, we can now think of some examples of ‘data ethical governance’ that may help shape public policy proposals and activities that specifically address human involvement and the level of autonomy of AI:
Legal frameworks that defend human powers. In his book *The Black Box Society* (2015), Pasquale proposed that legal frameworks are created for what he described as an “intelligible society” in which decision-making processes are always intelligible to all humans involved on a technical, organisational as well as societal level. This will require what he referred to as “humanising processes” (Pasquale, 2015, p. 198). That is, the establishment of company and policymaking practices that embed “human judgement” and involvement in automated decision-making processes (Pasquale, 2015, p. 197). He provided some implementable examples, such as human experts informing policymakers when their understanding of a technology is incomplete (Pasquale, 2015, p. 197). However, as he later proposed and elaborated in his 2020 book on a set of new laws of robotics, what we really need is a general framework for defending “human expertise in the age of AI” (Pasquale, 2020). We may think here of examples from concrete legislative frameworks developed in Europe in the 2010s; for instance, as I have mentioned earlier, the legal provision of the GDPR article 22 on automated individual decision-making and profiling (Regulation (EU) 2016/679) that basically aims to ensure the ‘human factor’ in such systems. In 2020, the European Commission furthermore published proposals for a Data Governance Act (DGA), a Digital Service Act (DSA) and a Digital Markets Act (DMA) with the overall objective to harness the power of large online platforms or what was referred to as the “gate keepers”. The Data Governance Act, for example, covered a call for investing in and supporting the development of trusted “data intermediaries” (that is; data trusts and stewardship models) to balance data asymmetries between large big data online platforms and individuals. The DSA on the other hand emphasised safeguards in regards to automated content moderation and data access to enable external audits and risk assessment of large online platforms’ AI systems.

Bottom-up governance approaches shaping human involvement in AI systems. Regulatory frameworks such as the GDPR and the proposed

---


125
DGA, DSA and DMA work as top-down requirements on the development and adoption of AI. However, other bottom-up governance approaches may also shape human involvement in AI development and adoption. By way of example, public institutions can use public procurement strategically to encourage the development of AI systems with heightened levels of human involvement (Hasselbalch, Olsen & Tranberg, 2020). Governmental and intergovernmental investment schemes can support start-ups that differentiate themselves on the global market with ethics-by-design innovation, products and services. In the micro context of design and development, engineers and developers need tools and methods to include human involvement components in their work with algorithms and machine learning models, such as “human in the loop” methodologies (Zanzotto, 2019), techniques for explainability and traceability (Gilpin et al., 2018), anonymisation techniques (Augusto et al., 2019), verification and validation and risk assessment tools (Menzies & Pecheur, 2005). Here, educational programmes can be implemented to increase developer competences and awareness, and shared technical engineering standards, can be supported.

In conclusion, as these examples illustrate, the ‘data ethical governance’ of AI is not simply a matter of discovering and harnessing the moral agency and ethical implications in the very design of AI. It means encompassing the entire chain of the distributed moral agencies of human and nonhuman actors.

4.11 Data Ethical Implications of AI

Now, a ‘data ethics of power’ suitable for AISTI governance needs an approach that encompasses the special power dynamics of AISTIs. Here, building on Part I: Power & Big Data, I therefore want to conclude this chapter by considering the special data ethical implications of AISTIs.

The computer hardware and software of the AI systems that store and handle big data are embedded in society and, like bridges, streets, parks, railroads and airways, they form our spatial environment, although they are different from traditional infrastructures. Roads and bridges, for instance, form the basic material architectonics of society and thus provide or limit access to places, but they are passive, so to speak, when mapping and
expressing human motives, morals and social laws. AISTIs transform the very objective material qualities of space. Quite literally they transform space into interconnected digital data. GeoAI (Krzysztof et al, 2020), for example, is a term used to describe the integration of AI systems and geography based on the analysis of data that contains georeferenced information ("GPS trajectories, remote sensing images, location-based social media, spatial footprints of buildings, roads, and parcels, global elevation data, land use and land cover data, population distribution, and so forth"; Hu et al., 2019, p. 2).

However, AISTIs are not just digital data extensions of material space. Recalling Lapenta’s (2011) depiction of the 21st Century “GeoMedia”, AISTIs also lock us in specific positions, providing or denying access based on the processing of personal data. They are mediating spaces that merge the human body, social and individual experiences, physical space and location into interoperable digital data, blurring their lines of separation when integrating them into the designed spatial architectures of a virtual infrastructure. (Lapenta, 2011) In this way, AISTIs function as the “new organisational and regulatory systems” articulating and organising social interactions (Ibid., p. 21).

GeoMedia constitute our everyday life spaces merged with mediated data spaces. In the 2010s these counted navigation tools such as Google maps and other location-based services such as the ride-sharing services Uber and Lyft. Then, there was also the “Geo spatial intelligence” system Sentient, which was under development by U.S. intelligence programs. The idea was that it would work on satellite pictures of the world with time and location stamps integrating all data, and eventually it would enable instantaneous and omnipresent AI analysis and strategy development of the U.S. military and intelligence (Scoles, July 31, 2019).

AISTIs are active infrastructural practitioners. They sense, learn and act based on what they learn, and evolve autonomously or semi-autonomously based on their interconnected big data environments. With a component of autonomous decision-making and behaviour, they actively shape the space they occupy. Crucially, I wish to argue here that the AISTIs therefore also actively participate in transforming the structure of our ethical experiences and critical practices. That is, they constitute an ethical experience, so to speak. Lefebvre (1974/1992) describes the architectonics of a space as something that does not just ‘exist’ but that are experienced. They are
defined by a body’s movement and sensing of its borders and directions. He calls this a “...bodily lived experience...” (Lefebvre, 1974/1992, p. 40). That is; space is, as I also described in part 1, material, social, but importantly also lived and experienced by humans.

Along these lines, we may also consider AISTIs actively self-producing spaces that amplify our experiences of a specific scientific, ideological and aesthetic paradigm. That is, as the historian Paul N. Edwards expresses, they are modernity embodied in a lived reality of control and order (Edwards, 2002, p. 191). Think, for example, of Amazon’s automated tracking and termination system that was deployed in its warehouses in the 2010s. This was lived and experienced by Amazon workers who were pressed to “make rate” to pack hundreds of boxes per hour (Lecher, 25 April 2019): “Amazon’s system tracks the rates of each individual associate’s productivity and automatically generates any warnings or terminations regarding quality or productivity...”.

Or think of the experience of a student in the school district of Andra Pradesh, India, where Microsoft’s Azure Machine Learning was deployed to identify students at risk of dropping out of school. In 2018, the AI tool had identified 19,500 students at high risk based on predictive analyses of data such as gender, socioeconomic demographics, academic performance, school infrastructure and teacher skills (Surur, April 22, 2018).

To summarise, with big data we created a quantifiable, measurable space ready to act on. With AI we created an agent, the hands and the brains, of the big data infrastructure; however, it is a very particular kind of agent directed at managing future risks and potentials. Thus, AISTIs do not only produce space. They are “Destiny Machines” (Hasselbalch, 2015) that also act on the temporal categories of individuals and societies by utilising the past (big data repositories) for the sole purpose of controlling and streamlining the present and future into something useful within the boundaries of the system design (shaped by different interests in the data of the system).

In other words, AISTIs are spatial and temporal architectures interrelated via streams of big data. Their agency in the world is empowered by a seemingly random data interconnectedness that tells us where each of us needs to be in the larger scheme of things. But it is an agency without human intention. It has no interest in where each of us really wants to be or where

---

we collectively ought to be. AISTIs provide us with agencies like the agency of the ‘holistic’ detective Dirk Gently, who says “I may not have gone where I intended to go, but I think I have ended up where I needed to be”.\textsuperscript{16} Dirk unscrambles mysteries in ways completely incomprehensible to any human, even himself, by accepting his position in the universe’s interconnectedness of things that time and again take him on journeys across time and senseless places. He does so without questioning it, and without intention, because he knows that the universe will always place him exactly where he needs to be. By itself this is AISTIs’ core ethical challenge to human power and agency.

CHAPTER 5

Data Interests and Data Cultures
DURING THE COURSE OF this study, I met many AI developers working specifically on ‘ethical AI’ solutions. Working ethically with the data of AI, however, I learned had many different meanings. There were developers who were developing AI for social purposes and who would argue that not using data was ‘unethical’. There were also AI developers who would argue for privacy protection and accordingly for data minimisation in AI development. Granted they were part of the same ‘ethical AI’ movement, they still did not have a shared conceptual framework for resolving conflicts between the data interests in the technologies they were developing. They were, I understood, part of different data cultures with different values and interests in data organised in different – and sometimes clashing – conceptual maps of meaning. All in all, when speaking with developers of AI technologies, I understood how various interests in data are constantly at play in the developmental phase of a data-intensive technology.

By way of illustration, one of the AI developers I spoke to and who agreed to be quoted here explained to me as follows:

“AI is like very data hungry, so when we are building something, then we are like thinking what API is there that we can use for this, so what API Google provides, like Bing or Amazon, whatever, what API is there that I can take the data, and make use of this…” She worried about the data she was sharing with these cloud services: “we just say that this is covered by their privacy policy, but we don’t know exactly what are the clauses, I don’t know, to be honest”, and when in response to her concerns I queried her further about what she got out of using these platforms as an AI developer, she answered: “I have good hardware, so I can run things way faster than I would on my computer, so I get the speed”.

In this way, a design choice with a trade-off between the data interests of users, the cloud service providers and the AI model was made. Later, she recalled a case in which the data interests of a group of people within a specific occupation trumped a business client’s data interests (I have changed the occupation here to preserve the confidentiality of the developer and company in question):

“I remember there was one case, which (...) involved one occupation (...) so it was like there was florists involved, we wanted to use some information about their performance, but then there was the union of florists, so we had to negotiate with the union of florists in order to use some data about the
florists, (...). So they were like basically trying to protect the interests of florists because then you can kind of infer something about the performance of a particular florist which may, I mean we are not gonna use it, but maybe the company who is our client may be able to say something about the performance of the florist...”. When asked how a union got involved, she disclosed another design choice based on the negotiation and trade-off between data interests: “(...) we needed to talk to the union in order to use their data, but then in the end we didn’t use it”.

Why does a data designer make the choices she does when considering different interests in the data design of the technology she is developing? What role does her cultural environment play? In this chapter, I explore these micro settings in which the technical components of a sociotechnical infrastructure, such as AISTIs, take material form as expressions of interests orchestrated in cultural systems of meaning making. An understanding of this cultural organisation of meaning, I argue, may help the ‘ethical governance’ of the complexity of sociotechnical change.

In the previous chapters, I have examined the shape of sociotechnical change as a complex of human and nonhuman factors that are thrown together in design, adoption, use and governance. However, I argue in this chapter, not in an arbitrary fashion. Sociotechnical change is made, it has politics, it has cultures, meaning that it is not neutral or ‘natural’, but embedded with interests. It is not even it, but, as also illustrated in the example of the ‘ethical AI’ design movement, it is many. Many embedded interests; many taken-for-granted cultures; many views of the world, priorities and conceptual frameworks, in harmony or in conflict. This also means that the “technological momentum” (Hughes, 1983, 1987) that enables a consolidation of change represents a compromise between these multifaceted cultural interests. According to Hughes, each developmental phase of a sociotechnical system’s evolution produces a specific "culture of technology", which he also defines as the environment of the socio-technical system and the sum of invested interests (Hughes, 1983,1987).

The culture of a socio-technical system we may also consider a particular “normality” (Kuhn, 1970), a knowledge foundation, worldview and conceptual framework for the practices of the developers designing its technical components, the lawmakers governing its adoption in society and the citizens in whose everyday lives it is incorporated into. Culture is,
according to Hughes what shapes the system’s technological momentum. The fundamental idea here is that competing cultures, and accordingly interests, must convert to the dominant culture of the momentum or perish. (Hughes, 1983). As such, technological change is first and foremost a negotiation between interests and a question of power and very human interests in dominion. As Langdon Winner writes:

“Technological change expresses a panoply of human motives, not the least of which is the desire of some to have dominion over others...” (Winner, 1980, p. 124).

European Commission, Social Media Campaign, 2019.
INTERESTS AND TECHNOLOGY

Large sociotechnical systems, such as AISTIs, transform and evolve in a multifaceted mix of relations. Each component of this development expresses its own history of knowledge, priorities, needs and available means. By way of illustration, personal AI assistants are software programs that interact with individuals answering questions and performing different tasks for them. Based on various AI components, such as voice recognition and machine learning, they are designed to act on behalf of and respond to individual users in a personalised manner. In a way, one can also describe them as the individual’s representative in the technical design of AISTIs. But in what way do they represent individuals? A data protection law may set the context of legal requirements for the assistant’s data design; fields of scientific studies will frame its technical and social potential; a developer will design its pre-set goals and priorities; a company behind might request a data design fit for marketing purposes; the user of the AI assistant will shape its self-learning process with their personal data; and so on. All ‘actors’ arrive with pre-set requirements and priorities contributing to the shaping of this personalised AI system with legal, personal, social, economic and political interests as well as worldviews and knowledge frameworks as to what is possible, necessary, beautiful and good, problematic, known and desired. However, not all interests are equally met in its final design.

Examining up closely the micro cultural histories and interests invested in the individual components of a sociotechnical system, the more general patterns of sociotechnical development might appear arbitrary and at best disorganised and uncoordinated (Misa, 1988, 1992). If we look at the design of just one personal AI assistant, for example, it does not seem to be more than just that, an assistant that will help a user in a personalised manner in a specific context with a specific service. However, if we examine it in terms of the general patterns of evolution of the sociotechnical system it belongs to, we see that all components of the system, including this personal AI assistant, most often have a common direction. They belong to a map of shared systems of knowledge and meaning making in which conflicts between different interests are resolved by negotiation or, what is most often the case, domination that will then characterise their technological momentum. (Hughes, 1983, 1987, Edwards, 2020, Misa, 1988, 1992)

What is important to understand here is that a technological momentum is
not formed within one single societal sector or community of stakeholders by, for instance, the economic actors that invest in the technologies developed. As Hughes describes, while the “mass” of a technological system is for instance influenced by the financial investment in machines, devices and structures, financial investment will not move the momentum forward alone (Hughes, 1983, p. 15). A technological momentum requires the efforts of all types of societal actors, from entrepreneurs and inventors to professional societies, organisations, businesses and governmental and educational institutions. This is also what he refers to as “...a supportive culture, or context...” (Hughes, 1983, p. 140).

Returning to the example of our personal AI assistant, we can finally see it as more than just a personal assistant in digital form. In fact, one can actually discern the general cultural shape, or “style” (Hughes, 1983), of the technological momentum that it is part of by examining the organisation of interests in its design. The personal AI assistants of the 2010s had been popularised within a technological momentum characterised by leading U.S. technology company actors as integrated components of mobiles, tablets or speakers with, for example, Amazon’s Alexa, Apple’s Siri, Google Now and Microsoft’s Cortana (Bonneau et al., 2018). The ‘big data technological momentum’ of the products and services, such as the personal AI assistants, of these large online platforms and technology companies were throughout the 2000s and early 2010s thriving on a big data “supportive culture” and “mindset” in public debate, among developers and users, and even to a certain extent in policies, such as some of the EU Digital Single Market strategies as well as individual member states’ digitalisation strategies. However, as I have illustrated throughout this study, this technological momentum was in the late 2010s also in crisis with rising critiques and revelations of in particular the data ethical implications of many of the big data products and services. Several critiques and concerns were for example raised against personal AI assistants that recognized voices and stored and processed their users’ personal data in more or less intrusive manners. In particular, concerns were presented regarding privacy and their prioritisation of the power and interests of the commercial actors behind versus that of individuals (Lynsky, 9th October 2019, September 4 2020; Chung et al., 2017; Maedche, et al., 2019).

In the early 2020s it had become clear that the distribution of interests embedded in the products, systems and services of the large online platforms and technology companies’ ‘big data technological momentum’ had to transform. A “supportive culture” of an alternative technological momentum in which the individual’s interest was prioritised was emerging in public discourse;
in standardisation (such the IEEE P7000s series); among entrepreneurs and inventors with different data designs and technologies (such as ‘data trusts’ and ‘personal information management systems’); in businesses and organisations with new forms of data governance and oversight (new ‘data stewardship models’); and also in legal frameworks (such as the EU DGA proposal) promoting different kinds of “data intermediaries”.

5.1 Stakeholder Interests

Now, we understand that interests are embedded in the evolution of a sociotechnical system and that these are negotiated, conflicts between them are resolved, and some interests are privileged over others when a system gains technological momentum. As a point of departure, we may consider interests in terms of different social groups or stakeholder groups. How do these different stakeholder groups contribute to the shaping of a sociotechnical system with different types of meaning creation? How do these groups actively formulate a system’s successes or failures, resolve conflicts and propose solutions? Bijker (1987) suggests that one of the things that technology communities do is developing and proposing “technological frames”, concepts and techniques to solve problems (Bijker, 1987, p. 168). These technological frames are invested with the interests of the communities that propose them (from micro scientific, personal needs, technological reasons or for macro-economic, political, and ideological reasons among others) and they shape the way in which critical problems are solved and, as a result, how the system will evolve.

A continuous appreciation of the position of the social actors that are affected by a technology is central to “ethical governance” (Rainey & Goujon, 2011). A core mechanism here is to actively include and ensure that a multiplicity of values and views are enabled and included in the processes of technology development and deliberation. As illustrated, in the context of the governance of the internet, the ‘multistakeholder approach’ was introduced during the WSIS process to ensure the inclusion of the interests of all stakeholders involved in the development of the internet and affected by the technology – from civil society groups and technical communities to industry and governments (Brousseau & Marzouki, 2012). However, the very act of including multiple stakeholder groups in a policy process of course does not ensure the fair balancing of their powers.

In the 2010s, many different stakeholder interest groups were represented
in official groups and bodies and were also explicitly vocal in public debate and politics in Europe concerning AISTIs: industry associations; consumers’ national and pan-European organisations, such as The European Consumer Organisation BEUC; digital rights NGOs, such as AlgorithmWatch, Privacy International, AccessNow, and European Digital Rights (EDRI); EU member states; national data protection agencies; EU bodies such as the European Commission’s different Directorate Generals; EU independent bodies such as the European Data Protection Supervisor (EDPS) and the European Institute of Innovation and Technology (EIT); national political parties; and European Parliament political groupings. Furthermore, other less organised stakeholder groups participated, such as various user groups, independent experts, activists, journalists, academics and individual companies. Still, the balance of the very contributions and activities of civil society groups and industry groups was often skewed by different levels of invested resources.

An illustration was the European Commission’s 2020 AI white paper (European Commission, I, 2020), which was indeed released for public consultation as part of a stakeholder consultation process, though the very contributions delivered came from 352 business and industry representatives and 160 civil society representatives. The very act of inclusion does not ensure representative inclusion. Another example to illustrate this point was the multi-stakeholder participation at the IGF held in a different country worldwide each year. All stakeholders could participate; nevertheless, participation was also limited by a lack of resources required for travel and accommodation.17

‘Ethical governance’ thus requires more than an intention and act to include; it requires an understanding of the needs and limitations of different groups in society and a conscious effort to meet those needs. Now, we may continue to explore the direction and shape of a technological momentum by tracing the interests of stakeholder groups in governance initiatives, and specifically their dominance in participation and inclusion of views in official legal and policy documents and statements. However, I want to propose that interests and values do not only pertain to distinct stakeholder interest groups that can be easily discerned and categorised in for example their participation in institutional activities and settings. As we

17 IGF does offer opportunities for funding that can be applied for by representatives from Least Developed Countries, Developing Countries or Transitional Economies (non-EU).
saw at the beginning of this chapter in my example of the interests influencing the work of the privacy-concerned AI developer, the complexity of their constitution is in no way easily disentangled. Interests and values invested in sociotechnical design are spread out between community “technological frames”, company goals, technological restraints, various legal quality standards, personal prejudices, needs and desires. With my point of departure in Hughes’ description of the technological momentum, I therefore argue here that ‘ethical governance’ requires an analytical view of the interests invested in sociotechnical change as a more complex set of factors that come together in shared knowledge frameworks and world views, which can be discerned with a view to the cultures that compete and also cut across the different stakeholder groups.

5.2 Interests in Social Contexts

As a starting point, interests have social contexts. They are shaped in economic and social contexts and accordingly are representative of structural power dynamics in society. As a classical sociological concept, interests are commonly considered determining factors for social action, or analytical categories for understanding societal developments (Spillman & Strand, 2013, p. 86). Interests can be discerned on a micro level in actions of individuals and stakeholder groups, or on a macro level in political, ideological and economic action. These “Interest-oriented-actions” are positioned towards goals and pursued as such (Spillman & Strand, 2013, p. 98). That is, they have an active agent and an object. How these agents, their goals and their object are defined, and the level of freedom of the interests in question is the result of a complexity of factors.

“Agency-theory” considers how trade-offs between the different principals’ interests are made by their agents, who act on their behalf in society (Spillman & Strand, 2013, p. 91). To illustrate, consider the members of the different stakeholder groups of AISTIs’ development; they are, as previously illustrated, represented in public debate and politics by a range of bodies and organisations. However, a member of a stakeholder group does not necessarily share the same interests as all the other members of the agent that represents them in politics or public debate. Thus, an industry association’s lobbying activities during the negotiation of a data protection legal reform do not just represent ‘the industry’s’ interest. When lobbying
for the wording of a particular provision, the association represents the compromise between different members’ interests. As I have illustrated previously, in the 2010s businesses had different quality standards for the development of digital data technologies and therefore different interests in the provisions of the GDPR.

5.3 Interests and VSD

We may continue here, as I do in the third article of the study on data interests and AI, to reflect on the distributed agency of the AI of AISTIs. Because interests in society are increasingly represented, distributed and realised in the design of partially autonomous technological nonhuman agents, it is crucial that we examine technological design as an ethically ‘non-neutral’ agent of interests; that is, as an agent that represents a compromise between different interests in society. Here, we may use the VSD framework presented in the previous chapter to address the moral agency of computer design when reinforcing values of stakeholders and distributing the agency of these by design. In VSD, interests are associated with the values held by different stakeholders. These values may be reinforced or repressed by a computer technology’s design.

In the 1990s, Friedmann and Nissenbaum (1996) looked at different types of bias in the design of computer systems that would systematically support decision-making that unjustly benefitted or disadvantaged some stakeholder groups, minority groups or individuals more than others. Based on an analysis of concrete computer systems, they developed three categories to discern how bias was embedded in design:

“Pre-existing bias” comes from the outside of the computer system where they ‘live’ in social institutions, in personal biases or attitudes held by developers. They are embedded in a computer system by explicit conscious efforts or unconsciously by institutions or individuals (Friedman & Nissenbaum, 1996, p. 333).

“Technical bias” emerges from “technical constraints or technical considerations”, such as limitations in hardware or software; the use of an algorithm that, due to its context of application, does not treat all groups equally from imperfections in pseudorandom number generation that, for example, systematically favour those at the end of a database; or as we saw in the previous chapter, from classification systems being insufficient to represent all life in a nuanced and fully representative manner (Ibid., p. 334).
Finally, “emergent bias” appears in the very context in which a computer system is used due to changes in population or cultural values, such as when a computer interface is designed for one type of use but applied in a different context, in which another user with different needs may not be sufficiently supported by the interface (Ibid., p. 36).

In this way, Friedman and Nissenbaum recognised that bias in computer systems is not just a technical issue but also the result of a combination of a complexity of social, technical and even individual conscious or unconscious efforts, which all have an influence on the position and treatment of different values and interests by design. Crucially, they illustrated how a computer technology’s design represents a compromise between different values held by different stakeholder groups.

Thus, an aim of VSD is to develop analytical frameworks and methodologies to resolve conflicts between the needs and values of different stakeholders in the very design of a computer technology (Umbrello & De Bellis, 2018, Umbrello, 2019). In this way, the embedded conflicts of interests of a technology are addressed as ethical dilemmas or moral problems to solve in the design and practical application of computer technologies. The approach therefore also seeks to instrumentalise the values held by different stakeholder groups, bringing them directly into the design process (Umbrello, 2019, p. 3).

By way of illustration, in the context of the interests embedded in a digital data technology, with VSD one can consider how privacy as a value held by a stakeholder group, such as digital rights organisations, might be adversely affected by the specific design of a data-intensive technology. However, one can also suggest, as I mentioned previously, an alternative design that is deliberately designed with privacy-preserving components (such as a “privacy-by-design”; Cavoukian, 2009) that enhance privacy values.

In the late 2010s, the idea that computer systems may have embedded bias in their design and accordingly potentially produce discriminating support or replacement of human decision-making was illustrated in a range of real-life examples of biased applications of AI. A 2020 study of patients in Boston, USA, for example, revealed how an algorithm used to score the health status of patients waiting for a kidney transplant was – by design, with the inclusion of race as a category – assigning black people with healthier scores (Simonite, 26th October 2020). Another example was the Beauty.ai beauty contest judge, which was supposed to provide the world with the
ultimate measure of human beauty, but instead represented the sum of its training data by favouring light skin contestants. Finally, there was the facial recognition software in digital cameras that analysed pictures of people of Asian descent as blinking humans (Mehrabi et al, 2019).

While technology design is a key focus of VSD, VSD scholars have also increasingly extended their stakeholder values analysis to governance contexts in which technology design is negotiated. Steven Umbrello (2019)’s examination of the way in which stakeholder interests are negotiated in “AI coordination” (the stakeholder coordination involved in what he refers to as “beneficial AI” research and development) is illustrative of such an approach (Umbrello, 2019, p. 4). For example, he examined the multistakeholder policy process of the UK’s Select Committee on Artificial Intelligence, which was appointed by the UK government in 2017, to consider the economic, ethical and social implications of AI and provide recommendations. He did so by identifying specific values (data privacy, accessibility, responsibility, accountability, transparency, explainability, efficiency, consent, inclusivity, diversity, security and control) in the committee’s evidence reports, tracing them directly to the different stakeholder groups involved in the committee (academics, nonprofits, governmental bodies, and industry/for profits) and ranking their order of distribution in the reports produced (Umbrello, 2019, p. 7).

In two of the articles of this study, I illustrate the role of interests in sociotechnical design and change. In the third article, I make a case for a data interest analysis of AI that explores how different interests in data are empowered or disempowered in the design of AI. In the second article, I address this type of negotiation of interests in the data design of AI in a geopolitical context, tracing and explicating cultural positioning as an interest in the AI momentum with an investigation of the unfolding of the European AI policy agenda on Trustworthy AI during 2018–2019.

5.4 Data Interests

In the previous chapter, I considered the AI of AISTIs as above all complex data processing systems and data design that form a type of embodiment of order, and in part I of the study I described the distribution of power in the information architectures of the Big Data Society. The data of an AI technology can also be viewed as an essential resource for these architectures and accordingly the locus of societal interests.
In the third article of this study, I use the human-centric approach of the EU HLEG on AI’s Ethics Guidelines for Trustworthy AI as an applied ethics framework for reflecting on the negotiation of power and data interests in AI data design. In the article, I define a “data interest” as an intention or a motive that is transformed into specific properties of a data technology that arranges data in ways that support the agency of certain interests in the data stored, processed and analysed by the AI system. By understanding data interests in this way, I draw attention to the formulation of political, business and civil society motives and intentions regarding digital data.

Here, I propose that a data interest is a motion to act on data in order to satisfy specific needs, values or goals that concern first and foremost data as a resource. Examples of such interests are political interests in data, commercial interests in data, scientific interests in data, the technical AI model’s interest in data, and the individuals’ interest in personal data. All of these data interests, I argue, are intertwined in the design of AI, but also in governance activities that seek to shape the evolution of AISTIs.

In the following, I examine how interests come together in general knowledge frameworks and values-based worldviews with enough force to shape, with standardized practices, development and adoption, the technological momentum of a sociotechnical system, such as an AISTI. However, first let me provide two examples of data interests ‘at work’ in the very design and development of smart city AISTIs.

First, we have Barcelona, which was one of the first European smart city initiatives to implement a data-driven, smart city infrastructure. This consisted of an extensive IoT sensor network collecting data about, for instance, transportation, energy and air quality. It included a bicycle sharing system with 6,000 bicycles, wireless sensors underneath roads to guide drivers to available parking spots, a waste management system with smart data-collecting trash cans, and smart lighting with sensors detecting when lights are required, in addition to such initiatives as for example saving energy and reducing the heat generated by the old lamps (Heremobility, 2020). This is of course an immense network of data that also includes the data of people, sensing and acting on a mobile environment, and mostly something a person moving through the city would not see or feel. But it is something all big actors with an interest in data resources do see very clearly: commercial actors with an interest in using data to personalise, train and better their services, scientists with an interest in improving results with
data, state actors to make services and processes more efficient and control the city. There are many different interests in the data resources of the smart city AISTIs of Barcelona. The main risk here is that only a few interests of the most powerful are met in the very data design of the city. However, in 2015, the new mayor Ada Colau took the smart city initiative in a new direction together with the city’s Chief Digital Technology and Innovation Officer Francesca Bria. Their mission was to develop data infrastructures “for and by the people”. Today, Barcelona has what they call a digital transformation agenda that views “data as commons”, opening up data to help the city’s entrepreneurial ecosystem, including SMEs in the ICT sector, and empowering citizens with tools that allow them to selectively disclose the information they would like to share. (Heremobility, 2020) Barcelona, together with Amsterdam, is also a pilot city in the European DECODE project, which develops smart city initiatives and tools where citizens can choose how and with whom they share their data.\(^\text{18}\)

The second example of the role of data interests by design that I want to use here is the centralised City Brain AI data system developed by the Chinese tech giant Alibaba. It monitors every vehicle in the city of Hangzhou, China, and has helped reduce traffic jams greatly; however, it also does numerous other things. The system constantly monitors video footage of traffic, looking out for signs of collisions or accidents to alert the police. It combines data from the transportation bureau, public transportation systems, a mapping app and hundreds of thousands of cameras. In this way, accidents are not only automatically detected and responded to faster but also things such as illegal parking are tracked live (Beall, 30th May 2018). Of course, there are also interests in the centralised data system of CityBrain, just like there is in the data system of Barcelona. However, the key difference between the two smart cities is that the CityBrain AISTIs do not have citizen oversight or control baked into their design. The huge amount of data generated by the system is designed to meet the interests of first and foremost a few power actors, namely law enforcement, the Chinese government, and the private company Alibaba.

In the 2010s, different technological cultures with different priorities in regard to meeting different patterns of interests in data by design, like these

\(^{18}\) The DECODE project: https://decodeproject.eu
two examples of smart cities in different parts of the world, were competing on a global arena for a momentum in framing the practices that go into designing the AISTIs and BDSTIs of the era. Here I propose that we are critically aware of these interests in the very design of AI, to understand what kind of power we enforce when we design the components of data intensive sociotechnical infrastructures: democratic powers, monopolistic powers, authoritarian or totalitarian powers. Because this is what we do. We create, provide and distribute power by design.
CULTURE AND TECHNOLOGY

As described in previous sections, Hughes (1983) considered cultural environments a crucial component of the fourth phase of a sociotechnical system’s evolution in which it gains momentum. In fact, a technological momentum is, according to Hughes, created by the prevalence of a dominant culture. It is this very common force that brings together all the diverse factors of human, social and technical character to create a technological momentum: “Taken together, the organizations involved in the system can be spoken of as a system’s culture.” (Hughes, 1983, p. 15). In particular, Hughes identified “cultures of technology” as “contextual elements” for the growth of a technological system that arise from the inside of the system in the shape of values and ideas of engineers and systems builders, but also from the outside in the form of cultures of a regional power (Hughes, 1983, p. 363).

Thus far, I have used the term ‘culture’ in the way that Hughes uses it (and it is used in STS in general) to describe shared conceptual and material frameworks for the development of sociotechnical systems. Moreover, I have used the term ‘values’, with a VSD approach, to designate the ethical and moral dimension of developments as such. However, what does it actually mean when we say that policymakers or engineers, data practitioners and systems builders share and practice a cultural values-based framework? That is, a shared culture that is forceful enough to create a technological momentum and make a system grow and consolidate in society.

5.5 Cultural Values and Technological Style

With a VSD approach to technological development, the moral evaluation of what we consider ‘good’ and ‘ideal’ becomes part of the technical design process. Values are in fact “idealized qualities or conditions in the world that people find good” (Brey, 2010, p. 46). However, these values are not just the personal ideals of individuals working with the design of a technology; they are also intentionally advanced by various types of stakeholders with shared interests and shared cultures. Culture is the foundation of ethical evaluation,
and culture is shared. This is, for example, the case with the shared cultural emphasis in Europe on the individual’s ethical agency, which is reinforced in a legal framework for the protection and promotion of ‘individual rights’ (Ess, 2014, p. 196).

Hughes also describes different regions as distinct settings for different ways of conceptualising and designing technology. As such, culture is represented in the “technological styles” of different regions. Differences in “technological styles”, he argues, became particularly apparent in the 20th Century due to the increasing availability of “international pools of technology” (including, e.g., international trade, patent circulation, the migration of experts, technology transfer agreements, and other forms of knowledge exchange; Hughes, 1987, p. 69). Technological style is an “adaptation to environment” (Hughes, 1987, p. 68), the technological language so to speak, of the culture of the economic and social institutions involved, in which knowledge and practice are systemised and conceptualised. In this way, even ethical evaluation during technological development can be argued to be a product of the culture and interests involved.

The interest-driven construction of knowledge production and conceptual foundations characterise one segment of STS approaches to the role of culture in technological development. Steven Epstein (2008) examines how the concept of culture evolved in sociological approaches to science and technology in two periods, from examining culture inside of the institutions and scientific labs to studying its role in the outside world of adaption and consolidation. Early studies of cultures inside scientific institutions brought forward the notion of knowledge as a cultural product, and thus brought with them a focus on the competition of scientific actors in pursuit of interests within distinct cultural environments (Epstein, 2008, p.168). Scientific credibility and authority were therefore also considered “cultural resources” with an emphasis on the very negotiation processes in which claims are made within scientific institutions (Epstein, 2008, p. 168). Thus, the very systems and networks of meaning making that frame technological practice have been a key focus of cultural analysis. In this context, culture was later examined as a multiplicity with an emphasis on distinct scientific cultures with, for example, ethnographic studies in science labs (Epstein, 2008, p. 169). Sociological accounts of science and technology reaching outside institutions and scientific labs, Epstein argues, on the other hand identify culture in conceptions of material culture and politics created
by humans to organise human life (Epstein, 2008, p. 172), such as with the technological means and modes of modern state governance. In the outside world, cultural consequences are also examined in practices of boundary creation or classifications of the world, which take form as modes of social ordering and power distribution (Epstein, 2008, p. 173).

In STS, culture is in general terms related to the way we get to know things and the skills and resources we use to create a technology. In the book *Science as Practice and Culture*, Andrew Pickering (1992) defines culture in a footnote as a resource for doing scientific work:

“Throughout this essay, "culture" denotes the field of resources that scientists draw upon in their work, and "practice" refers to the acts of making (and unmaking) that they perform in that field. "Practice" thus has a temporal aspect that "culture" lacks, and the two terms should not be understood as synonyms for one another: a hammer, nails, and some planks of wood are not the same as the act of building a dog kennel - though a completed dog kennel might well function as a resource for future practice (training a dog, say).” (Pickering, 1992, p. 3).

Distinct ‘knowledge cultures’ or ‘technological cultures’ can also be described as environments with rules for a technology’s design and adoption in society. The sociologist Harry M. Collins, one of the key people behind the sociology of scientific science studies at the British Bath School, defined “cultural skills” as intents and purposes and sets of implicit socialised rules of action for the design of a technology (Collins, 1987, p. 344). He considered scientific skills as belonging to different “explicable” or “inexplicable categories”. There are the formal facts and rules, the ‘heuristics’ ('rules of thumbs') and the manual perceptual skills that are visible and may be explained, and then there are the cultural skills that he describes as the inexplicable or “hidden” components of technology development (Collins, 1987, p. 337). These are required in order to use and understand formal facts and rules, heuristics and manual skills to develop technology. However, they are silently shared within communities and only acquired by the ones within the same cultural community. Thus, to an outsider without the cultural skills set, a crucial framework is missing, which is why in cases where different communities come together, the cultural component must be explicated. He

---

19 I do recognise that culture is also a contested concept in STS, e.g., as represented in the debate between Callon & Latour, 1992 and Collins & Yearley, 1992.
explains this as follows:

“When we interact with those who are close to us in cultural terms, we make do with few explicit remarks, and these convey meaning because so much is shared at the outset. But, as cultural and contextual distance increases between communicators, the potential ambiguity of more and more messages becomes realized, and more needs to be made explicit and can be made explicit.” (Collins, 1987, p. 344).

Consequently, Collins looks at the way in which skills within the different categories may transform and transfer into other categories, such as when a rule of thumb becomes a fact and formal rule written down in a manual. Crucially, he holds that the transformation of skills from one category to another involves changes in “cultural ambience” which is “enmeshed in wider social and political affairs” (Collins, 1987, p. 344).

Collins actually also uses AI expert systems to illustrate how implicit cultural skills may be transformed into explicit skills. With an AI expert system, all the skills of a human expert – the explicated formal facts and rules, as well as the inexplicable cultural skills – will have to be coded into the system for it to act ‘intelligently’. This may be done more easily, he argues, with human experts such as solicitors and medical specialists, who have skills that already rest on stores of codified information, rather than with specialists who rest on less organised expertise and cultural skills (Collins, 1987, p. 344).

This theory of cultural skills and skills transformation is an excellent example of the role of culture as a crucial component of technological development and change; that is, as the invisible and often taken-for-granted resources, skills and conceptual frameworks that are nevertheless crucial for technological development and change.

An AI system will only be as intelligent (and useful in its cultural setting) as its cultural design. Hence, failure to incorporate the invisible cultural component into an AISTI design will also imply malfunction in its consolidation in society. In a complementary manner, we may here additionally use a VSD approach to consider culture as a component of the very act of moral evaluation within the process of designing values into a technology, so to speak. Equally, failure to incorporate culturally sensitive moral evaluation into the very design of AISTIs will mean a clash with the ethical and moral evaluation of a given culture and society. Hence, as we can see, STS and VSD approaches place an equally strong emphasis on the role
of culture and shared cultural frameworks for sociotechnical development, although neither of these perspectives offers a conceptualisation of culture as such.

What is culture? How do we identify and discern the particularities of the cultural component and the cultural ethical evaluation component of technological change? In fact, how can we constructively understand the very qualities and characteristics of the different cultural systems of making sense of the world, which are held up against each other in power struggles and the interest negotiation of sociotechnical development and change as more than just conflicts and battles between systems? How do we constructively make sense of their compromises? I here turn to Cultural Studies to examine the concepts of culture and cultural values in more detail. I do this to understand how we may make the cultural shape and structure of the AISTI momentum of the late 2010s visible to guide it reflectively with a data ethical governance approach.

5.6 Culture and Power

At this point we understand that culture is a type of value system that brings together communities with shared conceptual frameworks and resources. Culture is also an active system in the sense that it constitutes specific priorities, goals and ways of organising the world that are actively imposed in society when practiced by, for example, engineers and represented in material things (like our technological systems). However, we can also crucially conclude based on the previous discussion that culture is constructed and invested with interests, and thus, the dominant cultures of specific communities and societies are only one view of the world. On those grounds, culture is like Bowker and Star’s classification systems (2000), or vice versa, the classification system is a cultural system of representation (I return to this point in more detail), practiced as if it was complete, but in effect a reduction of the world into one set of categories that do not represent the complete picture. Something is always left out. Something does not belong. That is, culture as a cultural system is never complete.

In 1958, one of the key founders of the British Cultural Studies tradition, the Marxist theorist Raymond Williams, famously defined culture as a “shape”, a set of “purposes” and “meanings” that are expressed “…in institutions, and in arts and learning” and in “ordinary” practice (Williams,
Culture, he argued, is more than just the refined, curated and selected art and literary works produced by one social class; it is also “popular” expressed in mundane everyday practice. It is “...a whole way of life...” (Williams, 1958/1993, p. 6). Culture consists of prescribed dominant meanings, but critically also their negotiations. Crucially, Williams argues that the meaning of culture is in “... active debate and amendment under the pressures of experience, contact and discovery ...” (Williams, 1958/1993, p. 6) and as such it is simultaneously “traditional” and “creative”. Hence, there are two sides to culture: “... the known meanings and directions, which its members are trained to...” and “...the new observations and meanings, which are offered and tested.” (Williams, 1958/1993, p. 6). In this perspective, we may also consider culture a site of power negotiation between the 'state of affairs’ and a potential for change.

Williams’ account of culture as “ordinary” and a “whole way of life” is essential to the British Cultural Studies tradition that emerged in the 1960s and 70s from the Birmingham school, with a specific focus on the study of popular and subcultures (Agger, 1992). Traditional elitist and exclusive conceptions of culture were replaced with studies of the culture of everyday life of a working class (Thompson, 1979/1963), and youth subcultures, introducing issues of race (Gilroy, 1987/2012; Hall, 1990/1994) and gender (McRobbie, 2000) representation and construction. That is, culture is not just one; it is multifaceted, it is institutionalised, but also mundane, subcultural – informally and formally created by and in interaction with people, including minority groups, and artefacts – and the meanings of these cultural relations are never stable. They are from the outset constructed systems of meaning making and therefore always up for contestation and social power negotiation.

Is there a ‘science’ for these cultural systems of sense-making? How can we make sense of the role of culture as more than just a black box that silently and invisibly defines practices and the design of our sociotechnical infrastructures? In Semiotics and Semiology, cultural practices, products and representations are investigated as components of systems of cultural meaning making that are interrelated and gain meaning through their systematic ordering. Rules of linguistics are applied to culture to formulate a ‘science of signs’, but it is not only the organisation of words in systems of signs and meaning that represent who we are and how we experience the world in culturally constructed systems. All types of systems of signs are
examined as language and accordingly cultural representation.

Roland Barthes, for instance, found cultural meaning in the mythologies of the French bourgeoisie expressed in everything from wrestling matches to advertisements for soaps (Barthes, 1957/1972). Myths, he argued, are “systems of signification” in which we make meaning of things in the world as elements of cultural discourse and order of specific moments in history: “...for a myth is a type of speech chosen by history: it cannot possibly evolve from the “nature “of things” (Barthes, 1982/2000, p. 94). As such, everything is, as he argued, culturally meaningful, including technological artefacts that have cultural systems of meaning making embodied in their very design. By way of example, toys of the 1950s in France, he illustrated, were designed as a micro cosmos of adult life and the roles and functions society encourages adults to perform: post offices, school, the army, medicine, and dolls that urinate and need to be taken care of by little girls. As such, they were cultural signs per se of the Bourgeois adult culture actively reproducing their patterns of social ordering (Barthes, 1957/1972, p. 53-55).

Along these lines, Stuart Hall defines culture as “systems of signification”, “conceptual maps” and “maps of meaning” (Hall, 1997). These are the conceptual systems in which we organise, classify, arrange and relate concepts with each other (Hall, 1997, p. 17). The cultural maps of meaning ensure that we understand each other and act coherently in a community when interpreting the world in the same manner. Cultures are, according to Hall (1997), social systems in which meaning is actively created and shared. Culture is not imposed on us from above. We actively practice, learn and belong to a culture. In this way, culture is also an active habit of sharing codes to communicate and make sense of the world, and these codes can be traced in cultural products that are interrelated in “systems of representation”. Specifically, shared codes are the keys to cultural systems of signification as they are the “stabilisers” of meaning (Hall, 1997, p. 21). Based on “unwritten agreement”, “conventions of representation” and “cultural ‘know how’”, the codes of culture are also the key to cultural belonging enabling us to “...function as culturally competent subjects.” (Hall, 1997, p. 22).

In his essay “Encoding/Decoding”, Hall (1980) describes processes of cultural meaning production as an interaction between a moment of coding and decoding cultural messaging. A television, for example, which was the key popular communications technology of the 1980s, he argues, is encoded with a cultural “…dominant cultural order...” (Hall, 1980, p. 123) with
“…preferred readings…” for their decoding receivers (Hall, 1980, p. 124). Although the process of meaning making is not “symmetrical”, according to Hall, and might even be distorted or misunderstood (Hall, 1980, p.119), the very moment of meaning making is embedded with the preferred readings that constitute a core component of the “…conditions of perception…” (Hall, 1980, p. 121). Whether reflectively and intentional or unintentional and habitual, these practices of encoding and decoding cultural systems are constitutive of culture that importantly reproduce the dominant cultural order (Hall, 1980, p. 123) Hall writes:

“These codes are the means by which power and ideology are made to signify in particular discourses. They refer signs to the ‘maps of meaning’ into which any culture is classified; and those ‘maps of social reality’ have the whole range of social meanings, practices, and usages, power and interest ‘written in’ to them.” (Hall, 1980, p. 123).

As we may recall, the VSD and STS scholars referenced in this study have also considered technologies ‘non-neutral’ cultural products embedded with ‘values’ and ‘politics’. Like Barthes’ toys and Hall’s television, Langdon Winner views technologies as linked with the power dynamics of a given society and asks:

“… Does this state of affairs derive from an unavoidable social response to intractable properties in the things themselves, or is it instead a pattern imposed independently by a governing body, ruling class, or some other social or cultural institution to further its own purposes?” (Winner, 1980, p. 131).

In their book Data Feminism (2020), the data scientists and feminists Catherine D’Ignazio and Laura F. Klein illustrate how through time, predominantly male data science cultures have been sustained in work environments where female data scientists’ work went unappreciated and unrewarded. These data science cultures, which the authors described as oppressive, are not only a gender struggle; they set the goals and priorities of the very data design in which power is distributed and where often minority groups’ interests are repressed; for example, when minority groups are either underrepresented in data used as the basis for decisions made on social benefits or when critical scientific medical analysis only benefits one privileged group, or on the other hand when a minority group is overrepresented in data that puts them at a disadvantage in society, such as data from specific city zones used for predictive policing.
In addition, in the very data science teams that develop the data design of the digital information architecture of our daily lives, D'Ignazio and Klein see an underrepresentation of minority groups. For instance, according to an AI Now report, women comprise only 15% of AI research staff at Facebook and 10% at Google (referenced in D'Ignazio & Klein, 2020, p. 27). These oppressive data science cultures are reflected in real-life experiences of minority groups working in data science and reflected in the data technology and design that have become an increasingly ubiquitous component of our social environment (D'Ignazio & Klein, 2020). Just as I also propose in this study, D'Ignazio and Klein therefore use the concept of ‘power’ as the axis for the injustices they see reflected in data science’s practices and data design:

“We use the term power to describe the current configuration of structural privilege and structural oppression, in which some groups experience unearned advantages - because various systems have been designed by people like them and work for people them – and other groups experience systematic disadvantages - because those same systems were not designed by them or with people like them in mind.” (D'Ignazio and Klein, 2020, p. 24).

A range of critical data studies are directly addressing the power dynamics of the environments and technological cultures that frame the practices and design of specifically AI and big data technological developments like these (O’Neil, 2016; Eubanks, 2018; Noble, 2018; D'Ignazio & Klein, 2020). However, critiques of the distribution of power in technological cultures of science and technology practice go further back. Since the late 1970s, a distinct research field counting feminist technoscience scholars, such as Judith Butler, Donna Harraway and Sandra Harding, has raised critiques of science and technology practices and knowledge in terms of the cultural gender power dynamics they reproduce and enforce (Åsberg & Lykke, 2010). Technology and science are here considered sites of dominion and identity struggle in which repressive categories of gender are fortified. Science is produced within scientific knowledge environments characterised by traditional gender roles, creating opportunities for some while repressing others. These are environments of repression that are then reinforced in the very technology created. The result is that existing power relations and dynamics in society are reinforced and many times even
amplified.

Understanding technology as a cultural product, and how technological practice is embedded in often repressive socially ordering cultural systems of meaning making that are lived and experienced by individuals, compels us to consider the cultural component of sociotechnical development as a specific object of our ethical scrutiny. This is also why the cultural systems of technology and technological practice per se are relevant as ethical problems that we should seek to solve with an applied ethics approach.

Following this, a core ethical concern we may have with AISTIs within a ‘data ethics of power’ applied approach involves their constitution as cultural systems of a type of social ordering, in which interests of dominant actors in society have the primary advantage while other minority interests are further disadvantaged. This is where we might also consider technology as a highly specific potential site of rebellion and freedom. Donna Harraway’s critique of the construction and positioning of gender through science and technology practice was, for example, voiced in her momentous 1985 Cyborg Manifesto, in which she imagined an alternative information architecture, a union between human and machine, based on “...socialist and feminist principles of design” that replace “the informatics of domination” (Harraway, 1985/2016, p. 28).

5.7 Data Cultures

Based on this understanding of the role of culture in technological development, I here wish to further elaborate the conceptualisation of data cultures in particular. That is, the technological cultures that frame data science and practice. With the previous depiction of the cultural systems of signification for practice in mind, I intend to address the culturally coded conceptual maps of data systems’ engineers, data scientists, designers, deployers, legislators and users. As I have argued throughout this study, these are not always shared even within specific stakeholder groups and communities, and they may even be in conflict; furthermore, they certainly are, as we saw previously, interrelated with societal power negotiation and struggle.

First of all, we may consider the very practices of data scientists and data designers to be framed within specific informal or institutionalised cultural systems of meaning making. Accordingly, the very practice of developing a
data system and design, can be understood as a cultural practice. At this point, I can without further explanation also argue that this type of work constitutes an active practice of ethical evaluation, or what Bowker refers to as “layering of values” into data infrastructures (Bowker, 2000, p. 643).

A data culture of a specific data scientist discipline is, as illustrated by Bowker (2000) in his examination of the cultures of managing the raw data of biodiversity databases, rooted in particular organisational histories, particular “temporalities” and “spatialities” (Bowker, 2000, p. 675); therefore, a data culture also constitutes active choices regarding what slice of the world is represented in the database. As such, we may as a first step define a data culture as constituted by cultural practices of ethical evaluation and choice when developing a data design (e.g., coding, labelling, managing, collecting and selecting data). The Information Studies scholars Amelia Acker and Tanya Clement explain this as follows:

“Understanding data cultures as underwritten by collections of data (as relata) means understanding data cultures as phenomenon shaped by ideas about the cultivation and production of data that reflect epistemologies about, for example, ordering, classification, and standards.” (Acker & Clement, 2019, p.3).

In the context of BDSTIs and AISTIs, it is relevant here to consider a paradigm shift in the way in which information science is applied when developing computer information systems, where what is taken for granted when designing a computer information system is uprooted and something different and new takes its place. This refers to a transformation of the knowledge and practice foundation of information and computer science from one cultural normality to another.

Computers are in essence information in a particular form. At face value, a computer is an information system that technically enables different types of data collection, sharing and processing for purposes that are supposed to be useful to humans. Information is processed, managed, modelled and classified according to mathematical formulas created by computer information scientists.

Like any science, the type of applied science of a computer scientist, or what I refer to here as a “data designer”, working on AI is to create a smoothly and efficiently working computer information infrastructure that senses, collects, stores and organises data and is trained and evolves on data to make ‘intelligent’ decisions. However, as I have illustrated, AISTIs are not just
neutral information processing systems. They have politics, values and a delegated moral agency that has ‘non-neutral’ social and ethical implications. This ‘non-neutrality’ is not in any way particular to an AISTI information system. Bowker and Star (2000) above all generally question the neutrality of the classifications of information and their standards, which they argue are integral to any working infrastructure. Classifications and standards permeate not only our digital lives, but have throughout human history also actively organised human relations with social and ethical implications for the people involved. From the classification of tuberculosis patients for purposes of incarceration in asylums to race classification during apartheid for purposes of segregation—these classifications and standards are not passive ways of effectively organising information (Bowker & Star, 2000). They actively create social and human life and therefore have an ethical dimension:

“We have a moral and ethical agenda in our querying of these systems. Each standard and each category valorizes some point of view and silences another. This is not inherently a bad thing - indeed it is inescapable. But it is an ethical choice, and as such it is dangerous - not bad, but dangerous.“ (Ibid., p. 5-6).

They argue that classifications and standards are related, but not the same. One is the practice of segmenting the world according to certain criteria, while the other is an institutionalisation of this practice of segmentation.

To start with, a classification is a “…set of boxes (metaphorical or literal) in to which things are put to then do some work - bureaucratic or knowledge production.” (Ibid., p.10). Classifications are consistent, have unique classificatory principles, and are mutually exclusive, which also means that you must either adhere entirely to one classification system or another for sorting information. There is, so to speak, no ‘outside’ of the classification system: “The system is complete.” (Ibid., p. 11). Of course, as Bowker and Star also argue, a perfect classification system is not possible in practice as there will always be ambiguity or disagreement as to whether an object belongs in a specific category. In addition, the information used to place this object within a specific category is in reality never complete. Human life, for example, does not fit easily into one classification system, and when it is reduced to one set of categories it will never represent a complete picture. In the previous chapter, I used Alpaydin’s example of human age, which is not easily placed in a ‘box’ by an expert system as we are not just ‘old’ or
‘young’, but are all ageing gradually (Alpaydin, 2016, p. 51). In the context of AISTIs, we also saw in the last chapter how one of the key drivers for the evolution of AI – from the 1970s expert systems to the increasingly autonomous big data machine learning systems of the 2010s – was overcoming the rigid limits of expert systems in representing their environments by building systems that could progressively perceive their nuances.

However, as was also illustrated in the last chapter, it is exactly this foundational ideal of the completeness of the computer algorithm’s classificatory work that creates the ethical and social implications when applied to human life. The computer algorithm is designed as a complete classification system and does not understand itself beyond this ideal of completeness. Yet, it is never complete and if deployed as such on human life it may exactly for this reason have grave ethical and social consequences (which was also one of Cathy O’Neil’s [2016] primary concerns).

For example, a predictive computer algorithm that makes risk assessments on the potential future criminal acts of an individual (such as, e.g., the COMPASS algorithm), may process different types of personal data regarding an individual to make these risk assessments. With a very simplistic example of an algorithm, this could be data about the individual’s location correlated with data on crime rates in different areas of a city (no algorithms deployed in these contexts are of course that simple). The individual might be living in an area with high levels of crime, which the computer algorithm then classifies as high risk. Because the computer algorithm’s principles for classification are always complete in theory, no other data about the individual is used to nuance the placement of this individual in a ‘high-risk’ category. If the algorithm is then deployed in a judicial system as a ‘complete system’, this may evidently have grave social consequences for individuals living in areas with high crime rates.

As a work practice among information scientists, classifications are embedded in the infrastructure of their working environments, invisible and shared in smaller communities. However, they may also become standardised and institutionalised, shared in more than just one community.

Bowker and Star describe Standards as: “...any set of agreed upon-rules for the production of (textual or material) objects.” (Bowker & Star, 2000, p.13). Standards are created to make things work smoothly together and they are enforced by, for example, legal bodies, the state or professional
organisations. Although no ‘natural law’ mandates their existence (and they therefore come into existence as the result of social negotiations among different stakeholders), they are heavily institutionalised and thus difficult to change (Bowker and Star, 2000, p. 14). Standards are, for example, key components of a well-functioning infrastructure and vice versa (Dunn, 2009). However, it is not only technical components that need to function in a standardized manner in order to work efficiently together with other components of a standard; the very work practices of people also need shared cultural systems of meaning making to function well. As Bowker argues elsewhere: “Working infrastructures standardize both people and machines.” (Bowker, 2005, p. 112).

In the context of AISTIs, examples of institutionalised technical standards that shape the work of an AISTI data designer are, for example, the International Organization for Standardization (ISO) systems requirements standards. ISO is an international standard-setting body with representatives from national standards organisations. It develops technical safety and quality standards for the development of products and systems, which it certifies to ensure they meet the requirements of the standards. ISO standards are internationally recognized, shared and developed. In this way, they ensure consistency in the design of a product or system, that the safety requirements are met, and that they are compatible with other products and systems that are compliant with the same standard. 20 The list of current technical standards for IT is long and diverse, spanning standards on IT security to information coding. 21 Although ISO standard certification is not a legal requirement, a certification of a system or a product does help ensure legal compliance.

In Europe, the GDPR presented a legal framework that in the late 2010s would be integrated into standards for the design of information technologies that process personally identifiable information. For example, the ISO/IEC 27701 published in 2019 specified requirements “for establishing, implementing, maintaining and continually improving a privacy-specific information security management system” with mapping to the GDPR. While the core purpose of the GDPR, “The processing of personal data should be designed to serve mankind.” (Regulation (EU)

20 Described on the ISO website: https://www.iso.org

21 Described on the ISO website “35 Information Technology”: https://www.iso.org/ics/35/x/
2016/679, p. 2), was not very different from the 1995 data protection directive that it replaced, the GDPR did have a stronger emphasis on the quality of the very technology design processes and practices set in place to ensure data protection and the rights of the individual.

To illustrate this with an example, the new provision on “data protection by design and default” (Regulation (EU) 2016/679, article 25) required that privacy and data protection were considered and specifically designed into an IT system from the start and not as an afterthought. This meant, for example, that what we have previously referred to as the “big data mindset” (Mayer-Schonberger, V., & Cukier, 2013) was challenged as one key task for a data designer would be to consider data minimisation as a core quality of the design of a data technology. Other technical design requirements included the pseudonymisation of personal data, creation of IT design with greater built-in transparency in regard to the functions and processing of personal data, and the ability for individuals to monitor the data processing.

Standards are centrally controlled and maintained, but of course they are not unequivocally embedded in practice. They are understood, interpreted and transformed by practitioners; negotiated in institutional, social and cultural contexts; and deviations are even accepted (Bowker & Star, 2000, p. 13-15). For example, the GDPR does not provide technical specifications. It is not a technical standard; it presents the legal compliance framework for technical and organisational measures and has a twofold aim to protect the fundamental rights and freedoms of individuals while ensuring the free movement of personal data within the EU (Regulation (EU) 2016/679). Data designers may in fact technically implement this legal objective with a weight on either one or the other side of this aim. A data designer with a “big data mindset” will imaginably seek to fulfil the data protection legal provisions only for reasons of legal compliance, but not as a quality standard for the very design of the technology. However, for another type of data designer, data minimisation and individual privacy could be a quality design goal per se.

In the 2010s, a growing design and business movement was developing information technologies with privacy and data protection as quality criteria for their work. In our book *Data Ethics - The New Competitive Advantage* (2016), Pernille Tranberg and I described a number of these cases of data designers and companies that presented their interpretations of the quality criteria for personal data processing with declarations of independence,
manifestos and public statements describing privacy and data minimisation as a quality standard for their work (Hasselbalch & Tranberg, 2016, p. 91). For example, the data designer Aral Balkan told us:

“...It is possible to build systems where individuals have ownership and control of their own data, on their devices, instead of holding it in a cloud where a corporation has ownership and control”. (Balkan in Hasselbalch & Tranberg, 2016, p. 92).

Another CEO of a German toy company Vai Kai, whose main product was a set of Internet-connected wooden dolls, Matas Petrikas, considered his customers’ privacy the basis of all design and innovation decisions. Therefore, Vai Kai’s internet-connected dolls did not have a camera and microphone like most other such dolls on the market at that time. He told us that privacy by design was a quality standard for his IT product:

"We think about privacy as a value all the time. It is part of our conversation. I assume other companies would never have had the conversation we had during our development phase that led to the conscious decision not to include a microphone". (Hasselbalch & Tranberg, 2016, p. 98).

Formal technical standards are difficult to change, but as Bowker and Star remind us, they are not mandated by “natural law” (Bowker & Star, 2000, p. 14). They represent a dominant culture’s values-based quality criteria. Therefore, they are time and again changed and updated in contexts of social negotiation. These changes of the standards for data design practices will represent, as well as require, paradigm shifts and new normalities. New priorities are set, new guidelines for practices are created, and new “scientific imaginations” and “worldviews” – to use the expressions of Kuhn (1970) – emerge.

The regulatory reform of the data protection legal framework in Europe was a symptom of a paradigm shift in the cultural environment of the information computer scientist/data designer. Moreover, the “data ethical movement” among designers and companies that Tranberg and I described (2016) represented a shift as such.

In the late 2010s, ISO standards in information technological practice were increasingly also updated to reflect a new normality for the data designers’ work. For example, a range of new standards for the development of AI were developed and published by the ISO/IEC JTC 1/SC 42 committee on Artificial intelligence. Several of these were, when I examined them in
early 2020, concerned with the big data design of AI, and many were also specifically focusing on the social implications of AI, such as ISO/IEC AWI TR 24027 on “Bias in AI systems and AI aided decision making”, ISO/IEC PRF TR 24028 on an ”Overview of trustworthiness in artificial intelligence” and SO/IEC AWI TR 24368 on an ”Overview of ethical and societal concerns”.

Another standards setting organisation, which in the late 2010s created standards reflecting a new more data ethically reflective cultural environment of the AISTI data designer, was the IEEE Global Ethically Aligned Design for Autonomous Systems P7000 series of standards with standards such as P7002 on the “Data Privacy Process of AI”, P7006 on “Personal Data AI Agents” or the P7012 standard for “Machine Readable Personal Privacy Terms” and many more.

To sum up, classifications of information are never neutral; there is always a moral dimension to the work of a data designer who is the human that inscribes the “programming language” of the technology’s delegated moral agency (Latour, 1992). This is also why Bowker and Star argue that a new information science normality is needed; that is, a “...new kind of science...” that links the social contextual awareness of social science with computer and information science (Bowker & Star, 2000, p.31).

In the 2010s, changes in the way in which the technical components of BDSTIs and AISTIs were designed involved a paradigm shift in what was considered the normality in data and information knowledge and science practices. That is, a real change of direction of BDSTIs and AISTIs in society meant a fundamental change in ways of working with information, how it was collected, processed, stored, analysed and used, which meant a change in the scientific imagination that shaped these processes. One step was the gradual realisation of the incompleteness of the systems’ data design when applied to human life, which came into existence when new types of scientific social and humanistic sensitivities were included in computer and information science.

Thus, we may also argue that the very data design of a technology has

---

22 Standards by ISO/IEC JTC 1/SC 42 Artificial intelligence: https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0

23 The IEEE Ethics in Action Initiative: https://ethicsinaction.ieee.org
cultural properties that can be examined as culturally encoded systems of signification. AI is, for example, not just ‘coded’ data; it is data culture in code. In fact, the AI system’s data design, or any data design is *culture in action*, enacting the cultural values and systems of signification: “…the database itself will ultimately shape the world in its own image: it will be performative” (Bowker, 2000, p. 675). For example, as outlined by Collins (1987) in his description of cultural skills and AI, culture is in expert systems transformed into explicated categories, literally coded and in advanced self-learning systems even encoded within the systems when autonomous machine predictions and decisions are made.

That is, the cultural classification of the world is actively coded and produced within the system. However, ironically in the specific case of autonomous systems, it simultaneously also scraps the cultural moment of cultural meaning production between an encoded “preferred meaning” and the decoded negotiations of meaning where, for example, conflicts with alternative cultures would otherwise have emerged had the interpretation been made in a qualitative spatially and temporally situated context and by humans. Or we can say more generally that the dominant culture is reproduced without challenge when transformed into seemingly neutral technical “professional code” that nevertheless, as Hall reflects when describing the cultural practices of professional television broadcasters:

”... operates within the 'hegemony' of the dominant code. Indeed, it serves to reproduce the dominant definitions precisely by bracketing their hegemonic quality...” (Hall, 1980, p. 126).

This is also why we may place emphasis on what I refer to as critical cultural moments of meaning making as a necessarily human component that cannot be reproduced by a data design or system that simply would literally reproduce a dominant code. Like Searle’s (1980, 1997) argument against the concept of strong AI, we may maintain as well that AI simply will never be able to ‘understand’ like a human. In other words, an AI system is not capable of reproducing the very cultural moment of meaning production, or what I in this study have also referred to as the “critical cultural moment”.

In conclusion: *data is culture*. It is not a natural given; it is not a raw material that may exist in itself and by itself (Bowker, 2014). Data has no meaning outside the cultural system, the database or the data practice. The data in the age of big data is in fact “datafication” (Mayer-Schonberger & Cukier).
Information is not just something ‘out there’, but is made meaningful in contexts of interpretation, representation and meaning construction (Mai, 2019).

Thus, we may propose here that the very cultural moments of meaning creation (the ‘critical cultural moments’) in the development of data systems are of crucial importance for an applied ‘data ethics of power’ approach, and accordingly readjust traditional approaches to the conflicts of data interests in big data environments accordingly. I maintain that a data culture’s moral and ethical evaluation is the most human component of cultural data design practice, which therefore needs a privileged position in the very design and development of BDSTIs and AISTIS. These, what I prefer to call human ‘critical cultural moments’ of ethical evaluation and meaning making, are negotiated and formalised in for example standards and classification systems. Moreover, they are not stable cultural categories but rather are time and again challenged, and different values and quality criteria (e.g., quality value criteria for handling data and interests) are negotiated and re-established with legal reforms, such as the GDPR.

The human ‘critical cultural moments’ may also be implicitly embedded in data design practices and expressed informally when data designers are negotiating and making design choices in a micro design context. Shilton (2015), for example, studied the values expressed by an internet architecture engineering team – the Named Data Networking project – and created a taxonomy of types of values of the data designers associating them with different interests: “(1) those that respond to technical pressures and opportunities; (2) those focused on personal liberties; and (3) those influenced by an interest in the collective concerns of an information commons.” (Shilton, 2015, p. 8).

Thus, in addition to being expressed in technical standards or legal requirements for the very technical components of a data designers work, they could for example also be expressed in ethical codes of conduct for data designers (Stark & Hoffmann, 2019). However, as described, a data culture is not just something that is taught and learned, and most often it is not even written down in standards and codes; it is lived and practiced by individuals. We may therefore in this context also consider what Shannon Vallor describes as practices of moral self-cultivation or “cultivating the moral self”, which she links to shared cultures of moral values habitualised and practiced under specific favourable conditions (Vallor, 2016, p.63).
PART III

HUMAN POWER & DATA ETHICS

“Mankind lies groaning, half crushed beneath the weight of its own progress. Men do not sufficiently realize that their future is in their own hands.”

(Bergson, 1932/1977, p. 317)
ABOUT PART III

In the previous two principal parts of this study, I formulated a ‘data ethics of power’ as an ethics critically concerned with the negotiations between the powers and interests of different sociotechnical data cultures, their conflicts and compromises. To make sense of the material of a ‘data ethics of power’ in the 2010s, I have described the two sociotechnical structures of power, BDSTIs and AISTIs, with each their data cultures.

BDSTIs and AISTIs constitute two forms of power that ‘work’ in different dimensions of human reality and society. While the BDSTIs primarily act in space by transforming all into immobilised digital data, AISTIs also occupy time by acting on that data to actively shape the past and present in the image of the future. I propose that here, at the beginning of the 2020s, a core concern of a ‘data ethics of power’ should be with these AISTIs’ and BDSTIs’ constitution as cultural systems of a type of social ordering in which interests of dominant actors in society are spatialised and immobilised, and thus, more difficult to be critical of and renegotiate.

Now, we have finally arrived in this study at the special characteristics of human power and ethical agency. I believe it is at the core of any engineered or nonengineered act of ethical governance of sociotechnical change as well as at the heart of a ‘data ethics of power’. A ‘data ethics of power’ is ultimately concerned with the role of the human as an ethical being with a corresponding ethical responsibility for not only the human living being but also for life and being in general. We need this human approach as a guiding narrative for the governance of AISTIs and BDSTIs; that is, an approach that prioritises the human environment, human ethical agency and responsibility. As I will show in this last part of the study, I use this concept with much explicit devotion to human judgement, governance and critical situated experience as opposed to the moral agency of technological artefacts, which can only represent, reproduce, reinforce living things without experience and critical agency. To illustrate this with an example, the main difference that exists between AI bias and human bias is that an AI agent cannot and does not negotiate its own bias.

A ‘data ethics of power’ is concerned with making visible these power relations in order to point to design, business, policy, social and cultural processes that support a human-centric distribution of power. The approach
therefore implies an applied data ethics in design and governance that ensures the involvement of human life, experience and critical agency in the very data design, governance, use and implementation of sociotechnical data systems. However, this applied human approach, this type of human power, needs specific spatial and temporal conditions to flourish. I argue that the BDSTIs and AISTIs of beginning of the 21st Century not only materialise power dynamics but also challenge human ethical agency to negotiate and revolt against these powers in fundamental ways. That is, human ethical agency is in constant negotiation with the moral agency of BDSTIs and AISTIs.

This is why we urgently need to preserve the human ‘critical cultural moments’ and create the conditions for our spaces of negotiation. We do need to actively build alternative sociotechnical data infrastructures and systems that interact with human agency, power and ethics in a different way. This we can only do by addressing the new sociotechnical power structures for human agency and experience with a particular kind of reality of human being in mind, a type of ethical agency that I believe does not resemble, correspond with, live well, or flourish in the cracks of the two sociotechnical structures of power that I have outlined thus far.
CHAPTER 6
What is Data Ethics?
WHEN THE DATA PROTECTION legal reform was implemented in Europe in 2018, risks and implications of the big data era were already household items in the public news cycle. The Cambridge Analytica scandal was one such thorny news item effecting waves of uproar in European media, policy and beyond. It revealed a British consultancy firms’ social media and big data analytics methods, based on the machine learning analysis of the data of 87 million people worldwide, including 2.7 million Europeans, to influence democratic processes in the US and UK (Stupp, 6th April 2018). Expositions such as these – data leaks and hacks, algorithmic discrimination and data-based voter manipulation – kept public attention tuned in to the data ethical implications of online everyday life, politics and culture in the late 2010s.

In public policymaking, discussions regarding data ethics had therefore equally gained traction and numerous public policy initiatives under headlines such as ‘data ethics’, ‘Trustworthy AI’ and ‘ethical technology’ were created in member states and in European intergovernmental policy contexts. In particular data ethical implications were equated with the new structures of power of the sociotechnical big data systems.

As such, European policy and decision-makers were increasingly positioning themselves against a pervasive, opaque form of power embedded in BDSTIs and AISTIs dominated by ‘GAFA’, an acronym for the four leading U.S. big data technology companies Google, Apple, Facebook and Amazon. Thus, in late 2019 and early 2020 the first steps were made to implement a European ‘Trustworthy AI’ and ‘ethical technology’ policy agenda in the shape of European laws and cultural values. Examples include the publication of a strategy for taking back control over a European data space and regain data sovereignty (European Commission, H, 2020) and even a potential ban on facial recognition AI (Delcker & Smith-Meyer, January 16th 2020).

Then one morning we went into lockdown. A pandemic was sweeping across the planet, and the European Union and governments all over Europe, all over the world, were scrambling to control, mitigate and predict the evolution of the crisis with various means and modes of governance. This included the swift introduction and adoption of several data-based digital technology and/or AI-based solutions. In fact, Europe experienced an immense digitalisation and AI boost, such as smart working and education
platforms, telemedicine, contact tracing apps, big data-based algorithms to support diagnosis and epidemiological studies, personalised medicine, and care robots (Craglia et al., 2020). When contact tracing apps were being developed in member states all over Europe, the debate on privacy and the choice between a centralised and decentralised management of data became fierce. However, it did not go further than discussions about privacy. In actuality, the BDSTI power of Google and Apple was only cemented more when the companies blocked the development of several contact tracing apps in European member states (Hasselbalch & Tranberg, 20th of May 2020). At the same time, Europe saw an acceleration of technologies handling personal data prioritising safety and public health: drone surveillance, location tracking, biometric bracelets, facial recognition and crowd behaviour analysis (Craglia et al, 2020). One may even argue that in the face of the COVID 19 crisis, only some of the ‘data ethics’ concerns regarding the distribution of power in the age of big data that had matured in Europe over the previous decade remained, while others were set aside (Martens, 2020; Vesnic-Alujevic & Pignatelli, 2020).

What happened to the power of data ethics in 2020 when COVID 19 swept through Europe? How could some core concerns be swept aside with it and only some remain? In this last part of the study on Human Power & Data Ethics, I present a formative framework for a ‘data ethics of power’. My main proposition is that a ‘data ethics of power’ cannot be put aside, neither can it just be applied when considered useful. That is, we can formulate data ethics guidelines, principles and strategies, and we can even program artificial agents to act according to their rules. However, to truly ensure a human-centric distribution of power, data ethics has to become more than just a moral obligation, a set of programmed rules, it has to be ‘human’. What this means, exactly, I will describe in this last chapter and part of the study. Data ethics needs to take the form of culture, to become a cultural process, lived and practiced as a way of being in the world. As such, a ‘data ethics of power’ first and foremost addresses the cultural conditions and structures of power, rather than the sole value properties of technology design.

There are a few premises for a ‘data ethics of power’ that I must outline first. Some are based on the conclusions of the previous two parts of this study that dealt respectively with the power of BDSTIs and AISTIs. In essence I here considered BDSTIs and AISTIs as particular types of ethical problems and accordingly data ethics a response to these specific challenges.
I argued that BDSTIs and AISTIs are the structure and shape of power for human agency and experience. I described how a complex of culture and power is immobilised in BDSTIs and enacted in AISTIs. As such, moral agency is increasingly also a property of AISTI agency and external to human agency. This does not mean that we are not still involved. We design technology, we use it, we interpret it, we shape it in our own image and interest. However, the very agency of our ethical evaluation and agency in particular is increasingly externalised.

This constellation, I argue, is a critical concern of a ‘data ethics of power’, because only humans can have the type of critical ethical agency that a ‘data ethics of power’ requires. A ‘data ethics of power’ is a revolt against a “closed” exclusive society seeking an inclusive “open society” based on love without a specific interest (Bergson, 1932/1977; Lefebvre, 2013) and the multiplicity of culture, and culture as a whole (Williams, 1958/1993). Therefore, ‘critical cultural moments’ and ‘spaces of negotiation’ are required to challenge the immobilised dominant culture of data systems.

The BDSTI and AISTI power structures for human agency and experience that I described in parts I and II of this study are core ‘problems’ addressed by a ‘data ethics of power’. However, a delineation of problems does not answer the key question of this final chapter: What is data ethics? To answer this question, I therefore propose an answer to two quintessential subquestions: Why is a ‘data ethics of power’ important? How can a data ethics of power achieve the ‘good society’? The answers to these two questions constitute the formative framework for a ‘data ethics of power’ illustrated in Figure 1.

1. Why is a ‘data ethics of power’ important?

First, I want to understand why we need a ‘data ethics of power’. What ontology necessitates a ‘data ethics of power’? What are the premises of our being in the world, and accordingly from an ethics perspective, what constitutes a ‘good society’ and ‘being’? These questions lead to the first formative component of a ‘data ethics of power’, which is an ontology that I will describe with reference to the French vitalist philosopher Henri Bergson’s ‘process ontology’ and the development of this by Gilles Deleuze, another French vitalist philosopher.
Ontology: Data ethics is a way of being in the world. It is an ontology of process and movement, where life is only stable and fixed when represented in systems of meaning making (Bergson, 1903/1999). Agents act in the world with different capacities. Humans are one type of agent, while technological agents, such as AI, are another. They both have agency, but not the same, as there are fundamental differences (Searle, 1980, 1997; Smith, 2019; Amoore, 2020; Pasquale, 2020). Therefore, in sociotechnical environments there are also two different ethical potentials. An AI agent can indeed be said to have a rational intellect and also act with moral agency (as illustrated in chapter 4), but it does not have the human ability to “think movement” (Bergson, 1907/2001, p. 318); it does not have the “semantics” (Searle, 1980, 1997), “doubt” (Amoore, 2020), “judgment” (Smith, 2019) or even “expertise” (Pasquale, 2020), and it can therefore never be an ethical agent by itself. Event to act as a moral agent of human ethical agency, the AI agent needs human empowerment, which is created by ensuring the ‘critical cultural moments’ (see below for a reminder of what this means) in design and adoption. This human empowerment is in essence what I consider the human approach of a ‘data ethics of power’.

2. How can a ‘data ethics of power’ achieve a ‘good society’ (an open society)?

Subsequently, we need to attack the core ethical problem of our age; that is, the BDSTIs and AISTIs in which power, culture and moral agency are captured and stabilized and in which the essence of our being to evolve and recreate ourselves in a constant process (and accordingly an open society) are immobilised. These core problems of the Big Data Society indicate the second formative component, the action-oriented approach, which will create the conditions for the critical human ethical agency that is necessary to achieve an open society.

Practice: Data ethics is a form of critical applied ethics that explores the conditions of power in the sociotechnical systems of the Big Data Society to actively create and ensure (‘data ethical governance’) the ‘spaces of negotiation’ and ‘critical cultural moments’. Spaces of negotiation are spaces carved in society with a material presence in
which values and interests are exposed and negotiated. Their core objective is critique and negotiation. They are possible when ‘systems’ (material/immaterial and technological/cultural) clash and controversy arises. For example, in policy, spaces of negotiation are initiatives established to negotiate values and establish shared ethical frameworks. However, they are only viable in moments where specific conditions make critical value negotiation possible (Hughes, 1983, 1987; Moor, 1985). *Critical cultural moments* have special human characteristics. They emerge and are only possible when human memory and intuition are privileged and provided time and space to tinker. For example, in AI design and adoption, the critical cultural moments are constituted by the level and type of human involvement and prioritisation of human environments in the technical design and in the adoption of AI systems.
COMPONENT 1: WHY IS A DATA ETHICS OF POWER IMPORTANT?

When asking this question, I am in effect with my answer also making a statement about the ethical capacities of humans and nonhuman agents in sociotechnical data infrastructures and thus their respective status and relation. What do we assume about human beings, their environments, technologies and crafts? How does a ‘data ethics of power’ perceive the relation between technology and humans, human agency and technological agency? The answer that I will delineate in the following sections is centred around the human approach that parts from another set of approaches to AI in particular, and data ethics and ethics of AI in general, which describe human biology as an information processing system comparable to the data processing of a machine (Wiener, 2013/1948; Floridi, 1999, 2013; Bynum, 2010).

First of all, while I do acknowledge humans as a natural part of the physical world (like Searle also does, 1997), I also propose that humans and their nonhuman agents do not have the same ethical capacities and therefore neither should they have the same ethical responsibilities. That is, humans are not information “organisms” or “objects” (“inforgs”, Floridi, 1999) comparable to other nonhuman information agents in a general information processing environment. Human existence and being is something other than the reception of information, processing and giving back of information, and consequently, I therefore also consider a human-centric data ethics as something else than just action on data. Data ethics is a way of life, a glimpse of an open society that only humans can grasp and therefore be responsible for.

6.1 THE HUMAN(-CENTRIC) APPROACH

In this study, I claim a human approach, or what elsewhere is also often referred to as the human-centric or human-centered approach to the ethical challenges of a Big Data Society. As I will illustrate, I am not alone in my devotion to a critical human (-centric) approach.

Certainly, a human approach is not a novelty in theories on human
existence, society, science and the world of our technological artifacts in
general. Nor is it unique in more recent analyses of the specific ethical
implications of the Big Data Society. Specifically, the human-centric
approach had a revival in policy discourse in the early 2000s on the
Information Society (see, e.g., the use of the term “people-centred” in the
declaration of principles, WSIS, 2003\(^{24}\)), and in the 2010s AI and data policy
discourse, as I also illustrate in the articles of this study. The Council of
Europe’s convention on human rights and Biomedicine (the “Oviedo
Convention”), which entered into force in 1999, also formulates an approach
based on:

”Primacy of the human being. The interests and welfare of the human
being shall prevail over the sole interest of society or science.” (Council of
Europe, 1997, article 2).

As such, when finalising this study in 2020, the human-centric approach
was a common term not only in theory but also in public discourse. Still,
said approach was presented in the policy discourse of the late 2010s with
no common conceptualisation other than an emphasis on the special role
and status of people and the human being. As a stand-alone concept this
could therefore mean many things, as a matter of fact it also did in these
policy discourses.\(^{25}\) Some of these meanings could even be said to be
ethically problematic, as also indicated by Mark Coeckelbergh, another
member of the EU HLEG:

“A human-centric approach is at least nonobvious, if not problematic, in
light of philosophical discussions about the environment and other living
beings” (Coeckelbergh, 2020, p. 184).

Here, I therefore want to offer an explanation of what I prefer to refer to
as a human approach that, yes indeed, takes its point of departure in human
nature. However, it does not prioritise the wellbeing of the individual human
being only. Rather it emphasises the role of the human as an ethical being
with a corresponding ethical responsibility for not only the human living
being but also for life and being in general (I will return to this claim with
reference to Henri Bergson’s concept of “human morality”, Bergson,
1932/1977). I argue that this is also the conceptual foundation of one


\(^{25}\) See e.g. excerpts regarding the ‘human-centric’ principle of the AI frameworks of the EU, Australia,
Japan, Singapore and OECD here: https://ai.bsa.org/global-ai-principles-framework-comparison/
particular human-centric approach, specifically advocated in European policymaking (and beyond), which I propose is distinct from other concurrent human-centric policy and business agendas. Let us here explore in more detail what I mean by this:

The human-centric approach was in European policymaking first and foremost framed in a European fundamental rights framework and with reference to a ‘human-in-command’ or ‘human in the loop’ approach to the development of AI that supports and enhances human agency and decision-making. It was also explicated in technology and engineering standards aimed at designing human agency in data technologies, such as the IEEE P7006 standard for Personal Data Artificial Intelligence (AI) Agents. In this way we may associate, as I argue in the third article of this study on the data interests of AI, the human interest in the data of AI in practical terms with the involvement of human actors in the very data design, use and implementation of AI. In the AI HLEG’s ethics guidelines, the ‘human-centric approach’ is by way of example spelled out with particular attention to the interests of the individual human being as well as the “human-in-command” and “human agency and oversight” components in the design and conditions for the development of AI. However, we can also trace the human-centric approach in more macro societal calls for action, such as Pasquale’s (2015) counter description to the “black box society”, the “intelligible society” where decision-making processes are always intelligible to all humans involved on a technical, organisational as well as a societal level. The human-centric approach is here actualised in what he refers to as “humanizing processes” (Pasquale, 2015, p. 198), such as concrete legal frameworks, which for example require the establishment of company and policymaking practices that embed “human judgment” in decision-making processes (Pasquale, 2015, p. 197).

These are highly practical proposals for a human approach to the Big Data Society centred around the value and promotion of human involvement and agency. However, there is an extra layer of reflection to this. In part I of this study, I positioned a ‘data ethics of power’ in the context of a recent data

26 See eg. Requirement 1 of the AI HLEGs ethics guidelines’s assessment list (2020) or the European Parliament’s “A comprehensive European industrial policy on artificial intelligence and robotics” (12th of February 2019)

27 See: https://standards.ieee.org/project/7006.html
(re)evolution of the information society; that is, the evolution of the Big Data Society. A prevalent characteristic of the Big Data Society is that it is dictated by a transformation of all things into data formats (“datafication”) in order to “quantify the world” (Mayer-Schonberger & Cukier, 2013, p. 79). I also argued that the sociotechnical infrastructures of the Big Data Society, the BDSTIs and AISTIs, are not the manifestations of an arbitrary evolution, but can be viewed as expressions of societal negotiations between different cultures, interests and at their very core also worldviews and ontologies of the status and capacities of the human being and the role of data technology in society. With a postmodernist perspective we might even consider the sociotechnical infrastructures of the Big Data Society the materialisation of a prevailing ideology of the scientific practices of modernity to command nature and living things (Jameson, 1991, Harvey, 1990, Baumann & Lyon, 2013, Bauman, 1995, Edwards, 2002). The critical infrastructures of the Big Data Society can therefore be described as modernity embodied in what Paul N. Edwards describes as a “lived reality” (Edwards, 2002, p. 191) of control and order:

“To live within the multiple, interlocking infrastructures of modern societies is to know one’s place in gigantic systems that both enable and restrain us” (Edwards, 2002, p. 191).

With a human-centric perspective, we may think of this as an ethical problem per se. Gilles Deleuze famously described over-coded “Societies of Control” (Deleuze, 1992), which reduce people (“dividuals”) to a code marking their access and locking their bodies in specific positions (Deleuze, 1992, p. 5). In other words, here we can also position the human approach within a postmodernist movement to free the living/human being from the constraints of the practices of control embedded in the technological infrastructures of modernity, which simultaneously reduce the value of the human being (Frohmann, 2007, p. 63). As we saw previously in this study, this was also the core critique carried forward in the field of surveillance studies (Lyon, 2001, 2018, 2014, 2010) as well as by Spiekerman et al. (2017) in their “Anti-Transhumanist Manifesto”. The manifesto directly opposes a vision of the human as merely information objects no different to other information objects (nonhuman agents); a vision, which they among others describe as “an expression of the desire to control through calculation” (Spiekerman et al., 2017, p. 2). In this way we can also consider the human approach of a ‘data ethics of power’ a critical reflection on the power of
technological progress, the sociotechnical systems we build and imagine. However, to do so we need to first address the Bergsonian process ontology of a data ethics of power.

I propose Henri Bergson’s process ontology (Bergson, 1907/2001) and “human morality” (1932/1977), which I outline in the following section, to be the first formative component of a ‘data ethics of power’. I do this to provide a foundational understanding of the ethical capacities of humans and ‘nonhuman’ technological agents, such as AI, that make out our sociotechnical environment, and to propose an objective for data ethical action to create the conditions of an open society for humanity. Bergson’s critique also covers the kind of society that I presented in the last part of this study in which human life is immobilised and disempowered in data systems of social meaning making and representation. In other words, he raised an important critique of utilitarian approaches to the living, and therefore, I argue that his critique can also be used to illustrate the limits of the intellectual capacities that AISTIs possess. That is, as I have illustrated previously with reference to John R. Searle’s Chinese Room argument, a type of intellect that may only reproduce syntax, but never semantics (Searle, 1980, 1997). With this perspective, the human approach of a data ethics of power is first and foremost an acknowledgement of the specific ethical potentials and responsibilities of humans as opposed to the intellectual potentials replicated in the autonomous moral agency of AISTIs.

6.2 Henri Bergson’s Human Approach

Henri Bergson’s process ontology is one that essentially resists rationalist representations of reality as realities per se. They are representations that utilise the real for our own purposes, which is an ethical problem by itself, as he illustrated in his final book The Two Sources of Morality and Religion published in 1932, which directly deals with ethics and morality. Bergson raised his key concerns with the limits of an intellect solely guided by a utilitarian approach to society in the early 20th Century in the context of world wars. This is relevant to know, as his concerns also addressed and were informed by the experience of the severe real-life human implications of a particular approach to the living. Scientific and technical innovation was in war time invested and shaped by the conditions of war and the interests of enemies and allies, and he had seen some of the devastating human effects
of scientific progress. Morality, Bergson therefore argued, could only be set aside like this during a crisis, such as war time, because it was practiced as a social moral obligation, not lived as a human morality (Bergson, 1932/1977). I will here illustrate what this means.

Bergson famously illustrated his critique with reference to the human conceptualisation of time. Time is a human invention, he stated, or it is “nothing” (Bergson, 1907/2001). Humans have created the mechanical structure of clock time to measure, segment and organise the time of the individual to function in society (Bergson, 1889/2004). However, clock time is not real time. It is the representation of the evolving time that we can only perceive while living it. Or phrased differently, we have two options for approaching life and reality: one is to approach it from the outside rationally with our own ready-made concepts, while the other is to experience reality in its “creative evolution”, as “duration” (Bergson, 1907/2001, Bergson, 1889/2004). The latter is what Bergson argued can only be achieved with human intuition. While indeed useful in science to act on material things (“matter”), the utilitarian intellect also, he argued, provided little room for a living, moving reality (Bergson, 1907/2001, 1896/1991, 1889/2004) and thus also an “open” and “inclusive” society (Bergson, 1932/1977). As he put it:

“In vain we force the living into this or that one of our moulds. All the moulds crack. They are too narrow, above all too rigid, for what we try to put into them.” (Bergson, 1907/2001 p.viii).

The utilitarian intellect may also constitute a type of morality. In The Two Sources of Morality and Religion (Bergson, 1932/1977), he proposes that there are the two options for morality: a “social morality” that takes form like our invented clock time, and a “human morality” that takes form as the latter human evolving time (Bergson, 1932/1977, p. 35-36). Social morality is expressed as a moral obligation that can be applied, but also set aside in, for example, moments of crisis. One might therefore also argue that it is prone to an interest-driven ethics. It can be used for purposes that serve specific interests. The utilitarian intellect, however, cannot produce the kind of “human morality” that constitutes a way of being in the world that we do not put aside or apply when needed, but have as a “style” or “way of life” (Bergson, 1932/1977; Deleuze, 1986; Lefebvre, 2013). As such, Bergson also advocated a different ethical approach, a “human morality”, which is also what I in this study refer to as a “human approach”, to ensure the open inclusive society in which it is not the human being as such that is prioritised,
but rather a “human morality”.

6.3 Henri Bergson’s Process Ontology

Henri Bergson’s process ontology is as complex as the reality he seeks to describe. To do this justice, I therefore will describe it in detail here. Time is not only a metaphor in a Bergsonian ontology; it is also a philosophical approach and a method. Accordingly, the two types of time that I referred to previously, or what Bergson refers to as “multiplicities”, also each correspond to their philosophical approach (Bergson, 1889/2004, 1907/2001).

One type of “multiplicity” is abstract. It represents only spatialized time, the measure of time, not the temporal reality of time. It is a homogeneous time that is divided quantitatively, and accordingly changes in degree (changes spatial magnitude) when divided. Again, if we use the time metaphor, a clock represents this type of multiplicity. Clock time is not continuous, but is divided into the instants 1 to 12 and so on. It is, as Bergson argues, a false continuum of time because it is abstracted from actual movement (duration) and is determined by spatial quantities (it increases every hour). It is time in a spatial form, which fragments duration into “moments” and does not take account of what happens in the “interval” (Bergson, 1907/2001, p. 21).

The other type of “multiplicity” is “duration”/”real time”. Duration consists of multiple times that extend into each other like a “…flux of fleeting shades merging into each other…” (Bergson, 1907/2001, p.3). Duration is the “whole” constituted by many different rhythms, of which human consciousness is just one of many. Duration is a heterogeneous multiplicity, and similar to Albert Einstein’s relative time, it is divided qualitatively; that is, when divided the whole of the “moving zone” changes simultaneously (changes in kind, intensity). This kind of “multiplicity” embodies what Bergson prefers to call the “real” continuum of time: time in its temporal form, its “continuous” form (Bergson, 1907/2001, 1889/2004).

The spatial segmentation of “real time” (representation) into external homogeneous structures (“closed sets”), Bergson argues, is a utilization of the real. Nature and the living are controlled and utilized for practical purposes. He reminds us that this homogenous structure is in fact not the “real”, but rather an objectification of the real. As such, the homogeneous multiplicity
is an “impure” continuum of time created by human reason, because it implies a stable universe; that is, a state of ‘being’ where the ‘whole’ is given from the point of departure and each unit is aimed towards a predetermined point of closure. Reality is in truth a “moving zone” (Bergson, 1907/2001, p.3) or an “aggregate of images” (Bergson, 1896/1991, p. 18) in which the human agent is one of many equally privileged agents. We might here also refer to Gilles Deleuze and Felix Guattari’s (1980/2004) more famous and quoted interpretation of reality as “the field of immanence” (Deleuze & Guattari, 1980/2004). The field of immanence” has no point of departure and no ending, and thus, it is continuous and open:

“…not internal to the self, but neither does it come from an external self or nonself. Rather, it is the absolute Outside that knows no Selves because interior and exterior are equally part of the immanence of which they fused.” (Deleuze & Guattari, 1980/2004, p.173).

In this reality (“aggregate of images” or “field of immanence”), human conduct is, according to Bergson, limited when confined by a utilitarian intellectual approach. The utilitarian intellect will only grasp its own possible action upon other objects (Bergson, 1896/1991, p. 21) and what is apprehended is solely the “best illuminated point of a moving zone” (Bergson, 1907/2001, p. 3). It is a type of intellect that is shaped and limited, as described previously, by the aim to utilize the real (Bergson, 1903/1999). However, humans also have other potentials; we are only limited if considering this type of intellect our only available capacity:

”… all doctrines that deny to our intelligence the power of attaining the absolute. But because we fail to reconstruct the living reality with stiff ready-made concepts, it does not follow that we cannot grasp it in some other way” (Bergson, 1903/1999, p. 51).

The human mind is a composite of “intellect” and “intuition”. Intuition we can associate with the human’s ability to “think movement” (Bergson, 1907/2001, p.318), our situated experience informed by a qualitative time or human memory (Bergson, 1896/1991) and more than any agent within a moving reality, the human thus has a potential for accessing the real time of reality, “duration”, as only humans have the potential to perceive duration with intuition.

We can use here an AI system’s action on reality to illustrate a tension between a utilitarian intellect and a human intuition, and accordingly why it makes sense to distinguish between the different applications of the two in
human environments. In December 2019, the AI company BlueDot Inc. demonstrated the great potential of AI big data predictive analysis when it raised an early alarm regarding a looming pandemic after having applied AI analysis to news reports and airline ticket data. However, in 2020 the big promise of predictive AI models like this was challenged by a human environment and behaviours in a fundamentally altered and unpredictable shape due to lockdowns worldwide and a global crisis (McLeod, August 14th, 2020). An AI system assumes an ontology of immobility, or in other words, a predictable reality of things, including human environments, but if we take for granted Bergson’s process ontology, the very properties of human environments are unpredictable and mobile. The historical training data on human behaviour that had been shaping AI predictions up until late 2019 simply could not deal with a present 2020 and a future beyond with unpredictable properties.

These different forms of human qualities of mobility (the very unpredictability of the human ‘critical cultural moments’), we may argue are challenged by the pervasiveness and potential normativity of AI systems. For instance, when using AI systems in judicial systems, a core risk is that AI tools become AISTIs of a human judge’s decisions. The European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe’s Ethical Charter on the use of artificial intelligence in judicial systems and their environment (2018) refers to this as “Quantity-based norms” (one of the two uses of AI in judicial systems the Charter considers with ”the most extreme reservations”, the other being profiling of individuals in connection with criminal matters). This is when AI systems’ quantitative analysis of the content of decisions produced by all judges turns into a norm privileged over the qualitative judgement of an individual judge, thus locking, as the CEPEJ Charter describes it, ”... his future choice into the mass of these “precedents” (CEPEJ of the Council of Europe, 2018, p.67). Phrased differently, but with a similar concern, Professor of Political Geography Louise Amoore introduces “doubt” as the most human component of an ethical decision-making process, thus challenging the solidity of “doubtless” decisions that are the result of machine learning processes weighing potential futures against each other and making room only for one probability:

“With contemporary machine learning algorithms, doubt becomes transformed into malleable arrangement of weighted probabilities. Though
this arrangement of probabilities contains within it a multiplicity of doubts in the model, the algorithm nonetheless condenses this multiplicity to a single output. A decision is placed beyond doubt.” (Amoore, 2020, p.134)

Returning to the human approach of a ‘data ethics of power’, we can now qualify the previous proposition that this is above all an acknowledgement of the essential value of humans as ethical agents in a sociotechnical environment. In fact, the replication of a utilitarian intellect in nonhuman intelligent agents is a core ethical problem that we may address with a ‘data ethics of power’, and here, we can also challenge the idea that ‘intelligent’ nonhuman moral agents can also be ethical agents. In fact, we may argue with Bergson that ontologically speaking they are not ‘ethical beings’. Indeed, it is evident that a technical system that gains its ‘intelligence’ (learns, remembers and evolves) via data/the spatialization of real time, can only possess one type of human intelligence, which is the one represented by what Bergson refers to as “clock time”. That is, a data system is always spatialized time, taken out of its temporal context and immobilised to be utilized for the purpose of the system. This is why the first component of a ‘data ethics of power’ is also a recognition of data ethics as a human responsibility.

6.4 Intuition as Method

With Bergson’s process ontology of existence as movement, we are also presented with a method for a ‘data ethics of power’. To Bergson, the ideal philosophical action is “to think movement” (Bergson, 1907/2001, p.318), to become one with the object of analysis; that is, to constantly renegotiate stable meanings. Intuition as a form of perception has here a specific status (Bergson, 1896/1991, p.66, p.185, p.183). In his book Bergsonism (1966/1991), Gilles Deleuze refers to this as “Intuition as Method”. This is an approach always moving towards an undefined point in the future as a state of “becoming”. It is also the approach that enables, what I here refer to as the ethical agent, to place herself within the qualitative and temporal context of a critical problem or ethical dilemma and consider the conditions of the problem. In fact, it enables the ethical agent to discover the problems and pose the problems. The ethical agent is for that very reason also ideally a free ethical agent.

Therefore, a starting point for a ‘data ethics of power’ is to find and
disclose the problems and solutions that may be covered up by stable systems of meaning production (or what I referred to in part II of this study as dominant cultural systems or orders). As Deleuze describes it, solutions and problems are inseparable from the systems in which they exist, and this is also why they are not easily detectable. A first step is therefore to uncover “…the conditions of experience…” (Deleuze, 1966/1991, p.23) (or to use Stuart Hall’s term “conditions of perception”) that have shaped already posed problems and solutions, or what Deleuze also describes as the problem’s “means and terms of stating it” (Deleuze, 1966/1991, p. 15).

This also includes uncovering the “false problems” that within one (dominant) cultural system of meaning production might be perceived as disorder, but in another is a cultural order in its own right. To illustrate this, we could think of the act of cleaning data of an information processing technological system (e.g., a software, a database or an AI agent) according to a specific classification system. This might be done in an ‘orderly’ manner, strictly adhering to the rules of this classification system, but at the same time we may discover that the very mode of ordering, the classification system, is a critical problem in and of itself. It might have ways of classification that are biased by representing only one dominant group, while causing an ethical problem, for example, for a minority group (as also illustrated by Bowker & Star, 2000). Here, it is the very order of the system that is the problem, and the ‘disorder’ (an alternative cultural system) that is the solution. Accordingly, we may argue as a rule that all data ethics problems are unique, just as all data ethics ‘solutions’ are unique—but also that they are uniquely interrelated as components of cultural systems of meaning making.

Deleuze provides a modus operandi that we can also use for a ‘data ethics of power’. We do not start with the solution. We go back to the problem and consider how it is “made”, how it is “set up” (Deleuze, 1966/1991 p. 15). The very problem per se is the expression of specific power dynamics that will guide the very solutions we propose and in which we encage, so to speak, our ethical agency. Deleuze states: “True freedom lies in a power to decide, to constitute the problems themselves” (Deleuze, 1966/1991, p. 15). What kind of social reality does the problem posed present? Who or what does it want to solve problems for? Who has an interest in solving this specific problem? To solve a problem, we need to find the problem, to invent it, which includes understanding how the original problems were stated and to
“uncover” falsely stated problems (Deleuze, 1966/1991, p.15-19). Only then can we consider solutions.

By way of illustration, privacy may be considered a problem in data technology and business innovation if framed by a “big data mindset” (Mayer-Schonberger & Cukier, 2013) within a surveillance capitalist economy, and technical solutions may be formulated accordingly. Thus, in this view of the problems of the Big Data Society, we may justify the development of big data systems with open data access and tracking by default of technical components with little privacy protection and safeguards. However, if we go back to the formulation of this ‘privacy problem’, we will discover that privacy is in fact not a problem, and it might even be a solution in and of itself. Hence, we can also consider privacy a type of data technology and business innovation (Hasselbalch, 2013 B; Hasselbalch & Tranberg, 2016). Deleuze also describes the Bergsonian method as a “Struggle against illusion” (Deleuze, 1966/1991, p. 21), a type of discovery of the real by uncovering representations of the real, which is equated with the conditions of experience. Doing this can only be done qualitatively, which includes the recognition of our own position in space. We take up “volume in space” with our very bodies and we similarly fill time with memory that links the different instants we perceive (Deleuze, 1966/1991, p. 25). In other words, our position and immersion in what we want to understand, study or solve, is simultaneously a strength and a weakness. It conditions our experience, but our human intuition also empowers us to uncover these very conditions.

Intuition as method constitutes a temporal approach to a temporal reality; a dynamic process in constant evolution. Thus, we may also argue that a ‘data ethics of power’ does not have a material form – it is not a guideline, a set of principles, a law, an initiative, a manual – it is processual and for that reason it has forst and foremost temporality. This is a crucial proposition, as sociotechnical systems are also constituted in time. They have, as I have illustrated in part I of this study with reference to Hughes (1983, 1987) and Moor (1985), patterns that can be discerned on scales of time. That is, when considering a specific data technology design, for example, we do not only address it as a type of occupation of space with particular properties, such as a set of fixed values (‘good’ or ‘bad’) that we can equally instil by design with another set of fixed ‘good’ values. We consider it a sociotechnical process, imagined, built, adopted, governed, reinvented and so on.
In this view, we do not only ask to what material thing do we apply data ethics (e.g., to which technology design), we also need to ask when is data ethics possible? As I have illustrated throughout this study, the data ethics spaces of negotiation are possible in critical moments of social controversy, “...when all the moulds crack...” (Bergson, 1907/2001, p.viii), the moments of ethical reflection where cultures clash and we make implicit values and interests explicit. These moments I can now also argue are the human moments. That is, essentially the critical cultural moments are the entire point of a human approach that seeks to ensure they do occur in the design, adoption and governance of sociotechnical development.

The conditioning of human moments, I argue, is also why a ‘data ethics of power’ is crucial in the context of BDSTIs and AISTIs. A core ethical problem unique to the sociotechnical infrastructures of the Big Data Society is the immobilisation of human culture and accordingly critical cultural moments in big data systems. That is, how do we support human critical cultural moments in an opaque sociotechnical development, in which cultural clashes and human interpretation are reduced to an automatic big data process? This is an ethical problem that we must address with a ‘data ethics of power’.

6.5 Data Ethics as a Whole Way of Life and Culture

In the part II of this study, I described culture as conceptual systems of meaning making that bring together communities with shared conceptual frameworks and resources. Cultural systems are also active systems in the sense that they have specific priorities, goals and ways of organising the world that are actively imposed in society when practiced by engineers, for example, and represented in material things such as our technological systems. But crucially, we also understand that culture is constructed and invested with interests and that dominant cultures of specific communities and societies are only one view of the world.

Cultural studies were founded on a critique of such stable dominant cultural systems. Raymond Williams criticised traditional elitist definitions of culture that presume a stable social reality constituted by enduring values, prescribed meanings and states of being:

"It is stupid and arrogant to suppose that any of these meanings can in any way be prescribed; they are made by living, made and remade, in ways we
cannot know in advance.”

He continued, with specific reference to culture in England, though with a much broader meaning:

“...the only thing we can say about culture ...is that all channels of expression and communication should be cleared and open, so that the whole actual life, that we cannot know in advance, that we can know only in part even while it is being lived, may be brought to consciousness and meaning.” (Williams, 1958/1993, p.10).

According to Williams this “whole actual life” (Raymond William’s ‘real’) can only be made meaningful in a society constituted by creative open systems of knowledge. Culture is many, and importantly, these cultures are not just one stable way of life, they are not only extraordinary (elitist), but “ordinary” (Williams, 1958/1993, p. 6); that is, whole ways of life. Thus, he confronts dominant cultural systems with a conceptualisation of culture as something that may be challenged and rebelled against with alternative cultural systems of meaning making. Here, we may relate the Bergsonian process ontology of a ‘data ethics of power’ to William’s conception of a dynamic, creative and crucially inclusive culture. If anything, culture is not a given, it does not have a stable meaning; it is a state of becoming, in negotiation and contestation. Exactly for this reason, I argue that the cultural component of BDSTIs and AISTIs is a specific concern of a ‘data ethics of power’.

A ‘data ethics of power’ considers technology a cultural product and technological practice embedded in socially ordering cultural systems of meaning making. In sociotechnical systems such as BDSTIs and AISTIs, culture – as we also saw in part II of this study – is actively practiced and lived by individuals, and cultural codes are taken for granted as stable frameworks of meaning sustaining power for some, while repressing the freedom and agency of others. Therefore, we may also approach the very cultural systems of data technology and technological practice as data ethical problems that we should seek to solve.

Here, as I have illustrated in this study, the very act of cultural criticism is particularly challenged by increasingly autonomous moral agencies where dominant cultural classifications of the world are actively reproduced and activated. This is why a ‘data ethics of power’ first and foremost seeks to recreate and ensure the critical cultural moments in the development and adoption of AISTIs and BDSTIs, where culture is treated as creative, "whole"
and “many”, and where negotiations of cultural meaning are enabled and voices of alternative cultures are meaningful. This is also where we arrive at the ethics of a ‘data ethics of power’. If culture is not one and stable, and if it therefore cannot be grasped sufficiently and represented honestly by ready-made cultural concepts, then how can we ethically approach it?

In a conversation with Didier Eribon, when asked about ethics, Deleuze answered: “It’s the styles of life involved in everything that make us this or that” (Eribon, 1986). Ethics is not the same as a moral obligation, not just the representation of the good and the bad; it is everything we do, our “style” of living. In *The Two Sources of Morality and Religion*, Henri Bergson similarly distinguishes between the two approaches “social morality” and “human morality” (Bergson, 1932/1977, p. 35-36). As presented previously, social morality he describes as one that is imposed as a moral obligation or duty in society. Therefore, we do not experience it as our own, and in this respect it is a kind of morality that we can resist and set aside. Human morality, on the other hand, is a way of living, a type of ethical way of being in the world. It is part of us as human beings, not represented in symbolic systems but expressed intuitively as an emotion in practice:

“... if the atmosphere of the emotion is there, if I have breathed it in, if it has entered my being, I shall act in accordance with it, uplifted by it; not from constraint or necessity, but by virtue of an inclination which I should not want to resist” (Bergson, 1932/1977, p. 48).

It is this human morality that I in this study consider ‘ethical action’; that is, the very agency of a ‘data ethics of power’. It is expressed in very subtle ways, in constant negotiation and contestation with a dynamic complex environment. Articulated in the style of our actions and practices, the nuances of different “technological styles” (Hughes, 1983) and of different styles of governance. Social morality on the other hand does not have a style: it is non-negotiable, inscribed in rules, and in machines.

On these grounds, we may also conclude here that data ethics is realised as a transformation of culture/a way of life (Bergson, 1932/1977; Lefebvre, 2013). As we will see in the following, this is a particularly critical realisation if we want human rights for instance materialised in sociotechnical systems as more than just adherence to and compliance with rules (Lefebvre, 2013). This does not mean that there should not be laws and shared common frameworks in society. Not at all. All it means is that these very solutions to ethical problems are not data ethics. What a ‘data ethics of power’ does is to...
ensure a critique, negotiation and reflection that will always result in a compromise that in itself could pose new ethical problems. A such, data ethics never has a point of departure, nor does it have an end, but is constantly moving with its target.

6.6 The Open Society

Now, we must return to the first formative question of a ‘data ethics of power’: Why is data ethics important? A human-centric ‘data ethics of power’ is important because it enables an open society. Bergson (1932/1977) describes two types of society: the “open society” and the “closed society”. The open society is also a type of open universal “love” that has no interest, but is universally directed at the whole of humanity (Bergson, 1932/1977). That is, the open society is characterised by a truly universal independent love. It does not have a specific object (or interest). The Bergson specialist, Alexandre Lefebvre (2013) describes this in his analysis of Bergson’s depiction of the purpose and function of human rights as a way of caring for and relating to ourselves:

“The open soul overflows with love, but it is not for anything in particular. Not for one’s family or nation, certainly; but also not for humanity or nature or gods or the universe.“ (Lefebvre, 2013, p. 92).

The open society is therefore also a just society in that it does not depend on any particular content and does not have a particular interest (Lefebvre, 2013, p. 95). To illustrate what this means in practice, Lefebvre uses Jankelevitch’s (1967/2005) example of the man who walks down the street joyfully smiling at everyone, directing his smile at everyone he walks by, but at the same time not at anyone in particular, and summarizes open love as follows:

“...love is a disposition or a mood. It is a way of being in the world, rather than a direct attachment to any particular thing in it.” (Lefebvre, 2013, p. 93).

The closed society, on the other hand, has “boundaries”; it is based on “...preference, exclusion, and closure.” (Lefebvre, 2013, p. 88) and is expressed in “authority, hierarchy, and immobility” (Lefebvre, 2017, p. 90). The closed society is one in which love is always progressively directed towards an object (the family, the nation etc.), a kind of morality imposed as a duty within a given society. It is represented and symbolic. Love and morality in the closed society expresses a “closed tendency” in that it is
directed towards a specific object. That is, love has an interest. It is dedicated to a specific group. Bergson exemplifies the core problem with social morality in the example of war. As described previously, he asks how human rights can be set aside in war time, and answers, only because human rights are realised as a moral obligation towards a specific group formulating this exclusive tendency of moral obligations accordingly:

“Who can help seeing that social cohesion is largely due to the necessity for a community to protect itself against others, and that it is primarily as against all other men that we love the men with whom we live?” (Bergson, 1932/1977, p. 32).

Thus, we have here two different types of love that materialise in the world as two different types of moral practices. One type of practice is adaptable, open, inclusive and moving, while the other is firm, closed and exclusive. We might also argue here with reference to the way I have used concepts of ethical agency and moral agency in this study that while the first can be designated as ethical practice and agency, the other is only moral practice and agency.

Lefebvre argues that it is in the very description of the open society that human rights gain a central position in The Two Sources of Morality and Religion, because they are “the best-placed institution to overcome the closed tendency of society and morality.” (Lefebvre, 2013, p. 83). Human rights are indeed considered universal in that they are applicable to all of humanity, meaning that human beings in all societies must enjoy the same rights. However, Bergson worries that the true value of human rights is not realised due to the way in which they are implemented in society in practice as a social morality. Lefebvre also maintains that human rights are not well expressed in moral obligation, in fact the goal of human rights is in essence better aligned with a changing state of being; that is, it is a transformative goal. Their function is to change the mind and personality of individual human beings and also of the state. For example, they not only protect individuals but also do so by reviewing and reforming arbitrary national laws and practices. Moreover, human rights do not only take the form of an obligation and compliance to law; they are embedded in cultural practice (Lefebvre, 2013, p. 75-81).

To illustrate the difference between a human morality and a social morality, allow me to provide an example of the life of an ethical value and human right, again privacy, in the context of the evolution of BDSTIs in the
2010s in Europe. I will as a point of departure consider privacy a human-centric value that enables the kind of open society I have just described. This is an argument by itself that I will not address further as it is explored and well-argued elsewhere (e.g., by Cohen, 2013; Solove, 2001, 2008; Hasselbalch & Tranberg, 26th December 2016; Veliz, 2020 etc.). In Europe, the right to privacy is also established as a legal right in the GDPR and the Charter of Fundamental Rights, which is embedded in member state laws, and as such can be considered a well-established moral obligation in Europe.

Nevertheless, one of the most profound challenges in history to privacy on a global scale (and the right to “private and family life”) was posed in the very brief history of the internet by evolving methods of surveillance, tracking and automated electronic systems of retention and correlation of personal data. Following the mass surveillance revelations by Edward Snowden, the United Nations General Assembly in 2013 affirmed that the same rights that people have offline must also be protected online. This statement was based on the realisation that the power distribution and conditions of the Big Data Society were challenging not only the legal implementation of human rights, such as the right to privacy, but these new constellations of power were also enabling the questioning of the very justification of a human right such as the right to privacy.

That is, in the short period in which the internet became a central part of global society, privacy as a protection of and moral obligation towards the individual was increasingly held up against other interests with strong arguments for setting privacy rights of individuals aside. While ‘anonymity’, and thus ‘online privacy’, was described in the 1990s as a unique opportunity offered by the internet to experiment with identity (Turkle, 1997) and gender (Haraway, 1985/2016), and challenge under its protection established forms of power and constituted market models (Vinge, 1981/2001), the right to privacy in the shape of online anonymity was also associated with things such as aiding and effecting identity theft, trolling (Donath, 1999), bullying (Kowalski et al., 2008), terrorism, and illegal sharing of copyrighted material (Armstrong & Forde, 2003). At one point, the very concept of individual privacy was even deemed obsolete or “no longer a social norm” (Johnson, 11th January 2010). As such, privacy, as the legal scholar Julie E. Cohen states, “got a bad name for itself” (Cohen, 2013, p. 1904). Thus, public discourse on privacy transformed increasingly legitimising arguments against the right to privacy as well as privacy-invasive business and state practices.
Furthermore, the very experience of individual privacy as a cultural component was influenced. In 2014, Verner Leth, Rikke Frank Jørgensen and I conducted a number of focus group studies among Danish youth regarding their use of social media. Although these young people did recognize in practice their own need for privacy with various forms of identity management, when asked about it, they were found to have simultaneously resigned to the idea that to participate in social life with their peers, they would also have to accept signing off their right to privacy to the social media companies that were facilitating it (Jørgensen et al., 2013; Hasselbalch & Jørgensen, 2015). In fact, they had negotiated their own interest in privacy with that of a social media business model and accepted to set aside their right to privacy.

However, in the slipstream of a sweeping data protection legal reform in Europe, a renewed public policy focus and public awareness of the privacy implications of BDSTIs were gaining traction, and in 2020 it was therefore also one of the core concerns in the debate in Europe on contact tracing apps. Did this mean that debates on online privacy had now finally matured into a human morality, a way of life, or in other words a data ethical culture that ensured it as a core value in European technology development, practice, adoption and experience? Or, as I propose, did it not just illustrate yet again the application of a social morality with specific interests? Why was privacy, for example, only an interest in the debates on contact tracing apps while in other areas it seemed to be a lesser concern?

A ‘data ethics of power’ would here urge us to look beyond the public debate on technical privacy to the more subtle power struggles and interests behind it. In the case of contact tracing apps in Europe, these were expressed in the wrestling of one power actor (European member states) and another (Apple and Google) that can be argued to have, in part, diverted the attention from other data ethical implications of not only contact tracing apps but also other data-based technologies developed during the crisis (Hasselbalch & Tranberg, 2020). That is, with a ‘data ethics of power’, we can address the structural distribution of power in the Big Data Society and argue that the implementation of human rights, such as the right to privacy, in BDSTI and AISTI development in Europe, was still only expressed as a moral obligation in society that could be applied or set aside according to interests and dominant power. Based on Bergson’s scepticism towards the realisation of human rights as a moral obligation, or as a social morality, we can also ask,
is this why privacy can be applied or set aside in the sociotechnical reality of the Big Data Society? I argue that indeed it is. In 2020, while data ethics time and again was affirmed as a moral obligation in strategy documents, principles and guidelines, it was still not a culture and a way of life, nor a technological style and practice, that could not be set aside or applied according to corresponding power interests.

With this ontology, and accordingly data ethics, we can now move on to the second formative component of a ‘data ethics of power’, namely an action-oriented critical framework addressing the conditions of power in the age of big data. If ethics and morality are ‘styles’, ‘cultural practices’ and ways of life of each individual, then how can we claim that they can play any role in the context of governance in society? That is, is there an approach to help direct society in an ‘ethical way’ beyond imposing it as a moral obligation only?
COMPONENT 2: HOW CAN A DATA ETHICS OF POWER ACHIEVE THE ‘GOOD SOCIETY’?

If the distribution of power in the big data cultures, technologies and societies that I have described in the previous parts of this study truly are in disharmony with our being in the world, with the ‘open society’ (what I have just proposed is the ‘good society’ of a ‘data ethics of power’), what can we do in an attempt to achieve harmony? What role can a ‘data ethics of power’ play in this?

In our book (Hasselbalch & Tranberg, 2016), we described data ethics as a social movement of change and action:

“Across the globe, we’re seeing a data ethics paradigm shift take the shape of a social movement, a cultural shift and a technological and legal development that increasingly places the human at the centre” (Hasselbalch & Tranberg, 2016, p. 10).

Phrased differently in the late 2010s a ‘data ethics of power’ was expressed in society as a proactive agenda concerned with shifting societal power relations and interests in the age of big data.

In the previous chapters of this study, I have described the data ethics of this agenda as a human approach concerned with making visible the power relations embedded in the Big Data Society and their conditions in order to point to design, business, policy, social and cultural processes that support the human interest and power. Consequently, I have proposed that power and human ethical agency are the anchors to which we tie a ‘data ethics of power’. One fundamental concern of a ‘data ethics of power’ is the power conditions of a human morality being reduced to a social morality in data systems and processes, which thus also inhibits an open society/inclusive love without an interest. Hence, we may also refer to data ethics as a rebellion against the reductive character of data systems and power, against exclusive power, and therefore that data ethics is predominantly the voice of the minority, the underprivileged and disadvantaged of the sociotechnical data infrastructures.

Crucially, a ‘data ethics of power’ makes invisible power dynamics visible in their temporal qualitative micro, meso and macro social and cultural
contexts. I have therefore in this study also relied on an STS analytical approach to sociotechnical systems that considers these simultaneously shaping and shaped by society (Hughes, 1987, 1983; Bijker et. Al., 1987; Bijker & Law, 1992; Misa, 1988, 1992, 2009; Edwards, 2002; Harvey et al., 2017). Translated briefly to our ‘data ethics of power’ formative framework, this essentially means that data ethics encompasses the roles of technology, culture and society in shaping power structures of human agency and experience all together. Thus, a ‘data ethics of power’ analysis, as exemplified in the articles of this study, is constantly moving on different plateaus in an effort to encompass the one and the whole (the micro and the macro) at the same time (Misa, 2009, 1988, 1992; Edwards, 2002).

6.7 Data Ethics of Power as Practice

A ‘data ethics of power’ is a Bergsonian human morality in action. As such, it is also a form of governance. This statement might be perceived as a contradiction as we have just learned that a human morality cannot just be applied (neither can it be set aside) but is lived as a process and practice. Nonetheless, I want to propose here that the very conception of a human morality could play an essential role in governance.

A ‘data ethics of power’ is first and foremost a human approach that ensures human ethical agency in, and responsibility of, data cultures. To this end, let me here repeat what has been said before about the concept of ethical governance in part I of this study. I focus on the role of data ethics and specifically the human approach. This includes a reflection on the critical moments of sociotechnical development when values and interests are negotiated and explicated and ‘spaces of negotiation’ emerge. The technological momentum required for a large sociotechnical system to consolidate in society is not just an arbitrary composition of social, economic and cultural factors mixed together by an inexplicable will of nature (Hughes, 1983, 1987); it has ‘human power’ and thus it may be transformed into different modes of governance that will guide the direction, the values, knowledge, resources and skills that form the technological architecture of the system, its governance, adoption and reception in society. But how?

As described in Chapter 3, ethical governance is “multi-actor”, “reflexive”, “open-ended” (Hoffman et al., 2017) and aims at setting in motion “... processes, procedures, cultures and values designed to ensure the highest
standards of behaviour” (Winfield & Jirotka, 2018, p. 2). Building on this conception, I propose to include ‘data ethical governance’ that addresses the complexity of the Big Data Society in particular and that pursues to ensure human ethical agency and responsibility in BDSTI and AISTI development. ‘Data ethical governance’ seeks and finds unconventional critical problems within the conditions and qualitative reality of the Big Data Society and constructs and restates problems and their solutions accordingly. Critically, ‘data ethical governance’ questions conventional problems and solutions by inquiring how, when and why these problems and solutions are posed and created (who has an interest in the problems and solutions that we take for granted?) ‘Data ethical governance’ asks: Which composite of solutions will best address the context and conditions of ethical problems? And crucially; how do we ensure a data culture in which the status of the human being as an ethically responsible and critical agent is acknowledged and ensured? To be exact, a ‘data ethics of power’, plays a crucial role as a human-centric frame of reference in the governance of BDSTI and AISTI development.

In conclusion, I propose that there are two key acts of a ‘data ethics of power’. These are necessary to achieve the critical ‘data ethics spaces of negotiation’ in which the values and interests are negotiated, problems identified and constructed, cultural compromises are laid bare and directions are recentered on a human centric distribution of power (Figure 2):

**Make power and interests visible.** One act is a disclosive and analytical process; a critical applied ethics concerned with data interests in the cultures and power dynamics of concrete data technologies and systems, data design and practices in companies and organisations among engineers and users and in politics. Interests and power dynamics can be discerned with a micro level analysis of the very design of a data system (as I have illustrated in the third article of this study); they can be examined with a meso level analysis of, for instance, the construction of political strategies on AI and data or the constitution of multistakeholder groups, legal negotiation processes and so on; and they can be investigated with a macro level analysis of, for example, cultural paradigm shifts, power dynamics and global cultural patterns, as I have done in the first and second article of this study. I have in this study referred to several examples of critical applied ethics as this, such as investigations of data systems, critical data studies, studies of discourse, legal
studies and surveillance studies.

**Ensure ‘critical cultural moments’.** Another act is to ensure the human ‘critical cultural moments’ in sociotechnical development and adoption. I have previously argued that these ‘critical cultural moments’ have special human characteristics, meaning they are possible when human memory and intuition are privileged and provided time and space to tinker. Therefore, ensuring these moments is also essentially what the human approach is all about. In Chapter 1 of this study, I illustrated how an ‘openness’ to the human ‘critical cultural moments’ can be practically ensured on a micro, meso, and macro scale of time in for example data design and processes, in institutional and company practices, and in the moments in between crisis and consolidation of a larger sociotechnical system in society (Hughes, 1983, 1987, Moor, 1985).

![Diagram](image)

**Figure 2. Data Ethics of Power Practices**

In the end, with ‘data ethical governance’ much is up to more ‘untraditional’ forms of human governance; the developers and engineers of BDSTIs and AISTIs; the people in the companies and organisations; the people that educate and direct others from primary school to university and to the work place; the activists that reveal and fight against the injustice; the scientist that develops the methodology; the ones that deploy BDSTIs and AISTIs, procure them and adopt them. Nevertheless, there is also some critical ‘data ethical governance’ components of the traditional governance work of the policymakers in the Big Data Society. In Chapter 4 I referred to these as “legal frameworks that defend human powers” and “bottom-up governance approaches shaping human involvement in AI systems”. I believe that in combination, all of these processes and practices can reshape the Big Data Society into a Human Society.
CONCLUSION

"In vain we force the living into this or that one of our moulds. All the moulds crack. They are too narrow, above all too rigid, for what we try to put into them."

(Henri Bergson, 1907/2001, p. viii).

A FEW MONTHS BEFORE his death in 1941, Henri Bergson left his sick bed to register at a police station. After France had surrendered to Germany in 1940, people of Jewish descent were required to register at police stations as part of the Vichy government’s anti-Jewish laws that among other things prevented a Jewish person from taking public office; being a member of the press, a student, a doctor, or a lawyer; and having a business. The Vichy government had offered to excuse Bergson, who was of Jewish descent, from these antisemitic laws due to his status as an internationally renowned academic, but he had refused. At the policy station, completing his police form, Bergson wrote: “Academic. Philosopher. Nobel Prize winner. Jew.” (Martin, 1994/2014, Chapter 10).

80 years later in 2020, in another part of the world, Robert Julian-Borchak Williams was arrested in front of his house and brought to the Detroit police station where he was held overnight for a crime he had not committed. In fact, he had been wrongfully arrested based on an erroneous biased match from a facial recognition system used by the police. Facial recognition systems like this had at that time been used by police forces in the US for more than two decades. Deployed for surveillance of specific communities and to identify people for prosecution, these systems had time and again also been exposed to reinforce racial bias. Presented with a grainy picture of the identified criminal, a black man, like Mr. Williams himself, but clearly not him, his first reaction was to ask: “You think all black men look alike?” (Hill, August 3rd 2020).
The data systems that we create to make sense of, organise and control life and society have throughout human history always reinforced power dynamics—often with devastating consequences for the human life represented in and by these systems; however, they have also changed shape. Today, the transformation of all things into data as an effortless, costless and seamless extra – and most often invisible – layer of life and society is one variety, which I in this study have argued requires a particular reflection and awareness from us. The data systems in which respectively Henri Bergson was registered in 1940 and Mr. Williams in 2020 both clearly represented and reinforced ethnic bias in society, and both systems constituted devastating ethical implications for the people to which they were applied. However, there are subtle differences. For example, while Bergson did not choose the system, he chose the data: a tiny personal rebellion, but nevertheless his comment against a devastating data system of power. Mr. Williams’ data, on the other hand, was chosen for him. In fact, he was not even aware that he had been registered in the data system that was now used against him by the police. We are here in the early 2020s not challenged by a database and register of a dominant regime of power, we are submerged in sociotechnical data systems of power. This is why we need a ‘data ethics of power’ to make powers visible and create the human critical cultural moments that ensure data ethics spaces of negotiation in society.

In this study, I have formulated a ‘data ethics of power’ to address new digital databased configurations of space and time (what I call “BDSTIs” and “AISTIs”) and the sociotechnical power structures for human agency and experience.

Over the years that I was immersed in the internet governance and digital rights policy, industry and civil society communities and debates, I became growingly concerned with the sociotechnical structures of power of BDSTIs and AISTIs, the liquidity and invisibility of this power, and an uncritical public debate about it. At first it was like fighting with the most popular kid in school; we were the outsiders, the activists, who did not understand the awesomeness of this reckless kid and his shiny new tools. Today, the privacy activists’ role and critical showdowns with powerful tech giants have become acceptable in public opinion. Still, even when the ethical reflection and social awareness are present, I see how we often fail to assess the implications of what we do across our different cultures of interests. We are immersed
socially and technologically in sociotechnical and cultural structures of power that limit us in what we do and what we think we can do with technology.

However, all is not lost. I do believe that an alternative route exists. We just need a different approach, not necessarily the interest and stake proper, the true idea, the right solution or the good ethics. We need a process constituted by love without an object. We need human power and a human morality.

The Study’s Three Key Contributions

1. A Data Ethics of Power

The core contribution of this study is a formative framework for a ‘data ethics of power’ concerned with making visible the power relations embedded in big data and AI socio-technical infrastructures (BDSTIs and AISTIs) in order to point to design, business, policy, social and cultural processes that support a human-centric distribution of power.

In the mid 2010s, data ethics at first gained traction in public discourse as a term to address the general socially and ethically problematic sides of big data technologies and systems in addition to their challenges to privacy. In many ways data ethics came to represent the ‘good intention’ of primarily companies and states. Then, it became part of an ‘ethics by design’ applied ethics, a moral philosophy, with methodologies and practices designed to instil good human values into big data and AI systems.

In this study, I part from a concern with the morally ‘good’ and ‘bad’ of big data and AI and position data ethics in the context of theories and critical studies that consider cultural and social power dynamics the essence of our ethical concerns with big data and AI. Technologies are cultural products and technological practice is embedded in socially ordering cultural systems of meaning making that are lived and experienced by individuals. This is why cultural systems of sociotechnical change and practice per se are relevant as ethical problems that we should seek to solve with an applied ‘data ethics of power’ approach. Thus, I argue that we need to be ethically concerned with the constitution of BDSTIs and AISTIs as cultural systems of a type of social ordering, in which interests of dominant actors in society have the primary advantage while other minority interests are further
disadvantaged. Sociotechnical digital data systems are spatial architectures that reinforce and distribute power. They have data cultures that sustain power for some while repressing the freedom and agency of others, and they are the locus of different powerful interests: corporate, governmental and even scientific.

2. Data Ethical Governance

Being actively involved in the evolving public debates and policy debates on the ethical implications of emerging data technologies, I have sought with this study to provide an explanation as to why data ethics became the centre of these discussions in the mid 2010s. By asking and answering this question, I have outlined a governance role and function for data ethics in the context of sociotechnical change. As such I have proposed a common ground for the debates and negotiations that I am involved in based on a human morality and approach to the power structures of big data and AI socio-technical systems.

Thus, I wanted to illustrate what data ethics can do for human governance in the context of sociotechnical change. I therefore examined the public debates and policy agendas of the late 2010s on the BDSTI and AISTI infrastructures of the Big Data Society as one ‘in between’ phase of the four phases of the sociotechnical development of large technological systems (Hughes, 1983, 1987), in which different technological cultures and approaches compete to gain technological momentum. I illustrated a “battle of systems” (Hughes, 1983) in which technical, legal, cultural and social components of different approaches to big data and AI existed in “dialectical tension” (Hughes, 1983, p. 79). I proposed that this phase of controversy is a critical moment for ‘data ethical governance’ as it is when ethical reflection, problem solving and value negotiation takes centre stage, thus potentially forming the cultural shape of sociotechnical change.

I argue that human governance of big data and AI adoption and developments entails a critical awareness of the ethical compromises we make in this moment of controversy as they constitute the cultural compromises invested in the BDSTI and AISTI technological momentum for global consolidation. Accordingly, vigilance of the cultural powers invested in this is vital. I here urge that we enrich a culture of human power in the governance, design and adoption of BDSTIs and AISTIs.
3. A Human (-Centric) Approach

I delineate a ‘data ethics of power’ as a human approach. In the late 2010s, the human-centric approach was a term and a theme that emerged particularly in policy discourse on AI with no common conceptualisation other than an emphasis on the special role and status of humans. However, the AI HLEG’s ethics guidelines were framed in terms of a distinctive European agenda created to ensure the development of a European AI ecosystem. The human-centric approach was here therefore also grounded in a European fundamental rights legal framework and the European data protection legal reform, with an emphasis on the autonomy and dignity of the individual human being.

In this study, I provide a supplementary conceptualisation of the ‘human-centric approach’ to the development of BDSTIs and AISTIs. Here, I propose a human approach that is one concerned with the role of the human as an ethical being with a corresponding ethical responsibility, not only for the human living being but also for life and being in general.

I argue that this is best expressed in the creation of data ethics spaces of negotiation and critical cultural moments. The spaces of negotiation enable critique and negotiation, but they are only possible when ‘systems’ (material/immaterial and technological/cultural) clash and controversy arises. The critical cultural moments have special human characteristics and are possible when human memory and intuition are privileged and provided time and space to tinker. In practical terms this is expressed in a prioritisation of the human interest in the data of BDSTIs and AISTIs via the meaningful involvement of human actors in their very data design, use, governance and implementation.

The approach of a ‘data ethics of power’ is not just about humans—it is human. This essentially means that it cannot be put aside, neither can it just be applied when considered useful. We need to think of data ethics as a human morality rather than just a social morality (Bergson, 1932/1977). That is, we can formulate data ethics guidelines, principles and strategies, and we can even program artificial agents to act according to their rules. However, to truly ensure a human-centric distribution of power, data ethics must become more than just a moral obligation, a set of programmed rules. It must be human.
Henri Bergson provides an excellent illustration of what this means with his process ontology. It does not mean that humans are bestowed with divine gifts. It does not mean that we are non-natural extra earthly beings. All it means is that humans do not only have the same intellect as machines. Our additional philosophical capacity is an intuitive one (Bergson, 1896/1991; Deleuze, 1966/1991), the capacity to “think movement” (Bergson, 1907/2001, p. 318) by setting into motion our memory (Bergson, 1896/1991). Most importantly, he raised an essential critique of utilitarian approaches to the living, and by doing so, I would argue that he simultaneously provided us with a conceptual map to also understand the limits of the intellectual capacities of both humans and AI, but that only AI cannot exceed. By way of illustration, while AI software can be trained by processing the data of 346 Rembrandt paintings to successfully create a unique 3D printed image that looks like a Rembrandt painting, perhaps even much better than a human reproduction, it could not do so without Rembrandt. A human could not either. However, what a human could do is produce their own painting uniquely positioned in time and space.

This is also a good illustration of what a human morality can do as opposed to the social morality that machines reproduce, and this is where we reach a data ethics of power’s essential meta critique. Today, AISTIs have increasingly powerful autonomous agencies that represent, reproduce and analyse the living in data with a utilitarian intellectual approach that utilizes the past and present for the sole purpose of controlling the future. That is, the world is made controllable and deterministic, represented in data that are interrelated as causes and effects. Data systems such as these, I have argued, cannot implement a human morality. They are “clock time” in action with no being in the world in its temporal form. Spatialized time, taken out of a temporal context and immobilised to be utilized for the purpose of the system. This is the ethical dilemma of our time that also concerns the disharmony of the data structures of control and the human freedom, as Deleuze (1992) and others have argued.

Furthermore, these closed tendencies of AISTIs and also BDSTIs are a fundamental obstacle to an open inclusive society, as love is exclusively targeted at specific objects according to specific opaque interests. This means that we do not and cannot know who receives love and who does not receive love, and accordingly we cannot ensure that love is truly universal. This is why I have argued in this study that data ethics needs to take the form
of a human culture, to become a cultural process, lived and practiced as a way of being in the world. As such, rather than addressing the sole value properties of technology design, a ‘data ethics of power’ first and foremost addresses the cultural conditions and structures of power.

The Structure of the Study

I introduced the study with a depiction of its integrated methodological approach, which was informed by embedded research, applied ethics and STS. Thereafter, I moved on to the main body of the study, which I unfolded in three main parts. In the first two parts, I explored the power structures for human agency and experience of BDSTIs and AISTIs. I described these as not only a critical concern of a ‘data ethics of power’ but also of ‘data ethical governance’ in general. In the last part, I delineated a ‘data ethics of power’ as a human power.

Conclusions Part I: Power & Big Data

In the first part of this study, I unfolded my discussion on power and big data in two chapters:

In Chapter 2, I delineated the special characteristics of the two sociotechnical infrastructures (BDSTIs and AISTIs) that a ‘data ethics of power’ addresses. I explored the spatial power architecture of the Big Data Society delineating in specific the first kind, the BDSTIs, in terms of their politics and invested imaginations and I addressed the ethical problems specific to the power architectures of the Big Data Society. Power, I illustrated, is concentrated, sustained and engineered by a few power actors, yet also increasingly self-sustained and evolving in cultures of use, design, governance and imagination and therefore difficult, but not impossible, to change. I suggested that understanding the “liquidity” of power (Baumann, 2000; Bauman & Haugard, 2008; Lyon, 2010; Bauman & Lyon, 2013, Castells, 2010) also means understanding the importance of a holistic ‘data ethical governance’ approach to the Big Data Society.

In Chapter 3, I considered the role of data ethics in the context of technological change. I here argued that we cannot only ask to what material things we apply data ethics (for example, to which technology design). We also need to ask when data ethics is possible. We need to do this because sociotechnical systems are constituted not only in terms of a transformation
of space, but moreover, because they have a temporality. They are sociotechnical processes – imagined, built, adopted, governed, reinvented, in crisis, and consolidated – and they must therefore also be approached as such in patterns of sociotechnical change that can be discerned on historical scales of time (Hughes, 1983, 1987, Edwards, 2002, Misa, 2009). In this way we can also find a temporal condition that enables ‘data ethical governance’. I illustrated how moments of explicit ethical reflection and value negotiation in the governance of sociotechnical systems emerge in intervals between crisis and consolidation, such as in the late 2010s when data ethics became a fix point in European policymaking. This period, I maintained, represented an ‘in between’ moment in which the critical sociotechnical problems of an existing data culture became visible and interests and values were explicated and negotiated. As a result, an explicitly cultural values-based governance position was formulated in Europe.

Conclusions Part II: Power & AI

In the second part of this study, I discussed power and AI in two chapters:

In Chapter 4, I examined the history and special characteristics of AISTIs, their ethical implications as well as the ethical theories that address these implications. The core objective of this chapter was to narrow the ‘data ethics of power’ considerations that were specific to AISTIs. I here argued that we need to take a closer look at the constructed relations between the distributed moral agency of active human and nonhuman actors. Social and ethical implications are in this perspective not just the result of a human intention, neither are they of nonhuman technological agency. Rather, they are consequences of a network of actions and competences distributed between these different actors. I explored AI as a new type of powerful technological agent that actively participates in moral decision-making processes (but cannot and should not take ethical responsibility), and therefore, is also increasingly shaping our ethical experiences. This is a core ethical dilemma of our age, I argue, and the reason we need a human approach to AISTIs with specific attention to the way in which humans are empowered and involved by the design and adoption of AI.

In Chapter 5, I returned to the question of how we can effect sociotechnical change with a holistic ethical governance approach. Previously, I had argued that although incredibly complex and not possible
to locate in one place, sociotechnical change is not arbitrary. It is made and can therefore be steered. Building on this, I now proposed that we consider culture and cultural practices as the basis of the transformation of sociotechnical systems and accordingly a focal point for ethical governance. Sociotechnical systems are developed and adopted in power struggles and compromises between cultural interests, or through the dominance of one cultural interest over others. In the 2010s, different cultural systems of meaning making, which I refer to as data cultures, invested with interests in the data of AISTIs and BDSTIs, were competing for a technological momentum in shaping the practices of developers, scientists, lawmakers and users. A ‘data ethics of power’ must address these very data cultures and their power conditions.

Conclusions Part III: Human Power & Data Ethics

The last chapter, Chapter 6, also constitutes part III of this study on human powers and data ethics. This is where I presented my formative framework for a ‘data ethics of power’, which consists of an ontology and a practice:

Firstly, I proposed that data ethics is a way of being in the world. It is an ontology of process and movement, where life is only stable and fixed when represented in systems of meaning making (Bergson, 1907/2001, 1889/2004). Agents act in the world with different capacities. Humans are one type of agent, while technological agents, such as AI, are another. I maintained that while an AI agent can act with moral agency, it is not an ‘ethical agent’. Even to be a moral agent, it must have human empowerment, which is created by ensuring – what I have argued throughout this study – human critical cultural moments in its design and adoption. This, human power/empowerment is, in essence, what I consider the human approach of a ‘data ethics of power’. Crucially, I here also argued that a human-centric ‘data ethics of power’ is important because it enables an inclusive, open society (Bergson, 1932/1977). Secondly, I delineated the action-oriented approach of a ‘data ethics of power’ that aims to create the conditions for the critical human ethical agency that can achieve an open society as such. I proposed that data ethics is a form of critical applied ethics that explores the conditions of power in the sociotechnical systems of the Big Data Society to actively create and ensure data ethics spaces of negotiation and critical cultural moments.
**What’s Next?**

During my research, I stumbled upon the delineation of the design of a mechanical ‘ethical governor’ of autonomous weapons. This would be a technical component created to process the stream of data that shape the agency of an AI warrior agent, a type of mechanical data control component built into the agents’ data design to ensure that it behaved within a set of prescribed ethical boundaries. (Arkin et al, 2009). The ethical governor would work by processing the data of the agent’s fields of action and permit or forbid actions based on this. There would, for example, be the data on humans, essential buildings and meaningful cultural sites, which the agent’s data design would transform into data streams of forbidden lethal action; then, there would be data on threats that could also include data on humans, as well as data threat scenarios that, in combination with other data on the ethical bounds of the ethical data governor, could be transformed into data torrents of permissible lethal action. This represents hundreds of years of military law and complex ethical boundary creation transformed into the data process of a data design.

Making ethical decisions on life, death, culture, bombs and lethal action is exceedingly complex, but ethical decision-making transformed into a data process appears to be less so. Today, this reduction of human complexity, I argue, is one of the key motivations for BDSTI and AISTI development and adoption: to make life, society and culture easier to handle; and to make those difficult ethical decisions we must make every day less cumbersome as individuals at home, work, school, and in our societies in hospital, during elections, in the welfare system, in the justice system and during times of crisis, such as war or pandemics. These are thorny human decisions, which we also time and again realise ourselves, or are told by other humans, we have made poorly or in ethically problematic ways.

However, this kind of self-conscious critique, that a data process does not have, is also exactly why we need to keep making these decisions ourselves. Think about an ethical evaluation of the most critical situation you can imagine, such as releasing a bomb during wartime, transformed into a data process without critical agency. It is terrifying. Nonetheless, that is precisely the kind of imagination concerning the reduction of complexity and human dilemmas of ethical decision-making that drives much of the development and adoption of BDSTIs and AISTIs today.

It is also this loss of human critical agency that I fear the most will take
form in our sociotechnical realities if we do not halt and redirect current BDSTI and AISTI developments. We do not seem to realise it because cutting out the human does not mean cutting out moral agency (making a decision, even a moral one, a data process certainly can always be designed to do). However, in this process, as I have tried to illustrate in this study, we cut out the kind of critical ethical human agency that is fundamental to our democratic societies and their institutions. Human critical agency is what we remove from the very configurations of our sociotechnical spatial architectures and from our societal imagination, our norms and cultures.

Now, we have some core tasks ahead of us to steer the change we need. These will all go into the development and consolidation of a highly complex socio-technical system that will be shaped by a multitude of cultural, economic and social factors; thus, I can of course only mention a few of the tasks here:

We need a new imagination, a different technological culture or data culture of human power, when building and adopting AISTIs and BDSTIs. We must stop trying to design our very human yearning for the perfect society, the efficient society and the just society into them. Humans, biology and societies are messy unpredictable things. They are anything but perfect; often they are unjust, and definitely worthy of our ethical concerns and critiques. However, big data systems or AI will not change that. Only humans can make real changes by ensuring the conditions and structures of power that enable human critical agency. What we can do, however, is to imagine BDSTIs and AISTIs as potentially incredibly useful tools that can support human critical and ethical decision-making with evidence; for example, apparatuses that can make a scientific analysis stronger, or instruments that indeed can help us make many processes more efficient.

We need technical components of sociotechnical infrastructures that enhance the critical agency of individuals. We absolutely must ensure meaningful human involvement, which we can do by carving out a place for individual human experience and agency in the sociotechnical fabric of BDSTIs and AISTIs. The first place to start, which particularly Europe seems to be investing in at the beginning of the 2020s, is the development of a data infrastructure that by default respects people’s privacy and empowers individuals. Today, the technical components of a data infrastructure as such go by many names, such as data trusts, personal data management systems
and personal data stores, and they take on many forms. Much is left to be explored both in terms of their basic functioning, interoperability and not the least their legal framework. Here, we also need to build in by design meaningful human in the loop/in command components, and we need data systems in which critical functions and criteria can be explained.

Moreover, we need people who understand big data and AI systems well enough to engage with them in a meaningful manner when they use them, procure them, create laws about them and build them. We need sociotechnical data literacy and education of children, educators, students and policymakers.

Furthermore, we need to update our legal frameworks to ensure the legal implementation of meaningful human control in the development and adoption of BDSTIs and AISTIs and to safeguard human expertise (Pasquale, 2020).

We need more critical data studies that discern the data interests in the cultures and power dynamics of specific data design, practices and data politics. I have in this study referred to several examples of brilliant investigative studies in these areas; however, most have been conducted in Western contexts, making visible their power dynamics and challenging the ethical problems and dilemmas of BDSTIs and AISTIs in these specific cultural spheres. It is time to voice the essential experiences of the people, communities and cultures that are often the most disadvantaged in the global big data structures of power. This is why we need more critical data studies that explore and emerge from these different cultural and socioeconomic contexts.

Last but not least, how about human ‘ethical governors’? What if we designed independent (I mean truly independent in terms of not only finance and interests, but also in terms of their cultural conceptual frameworks of meaning making/imagination and narratives), multi-expertise (not multi-stakeholder) ethics councils and bodies into the governance of AI and big data sociotechnical development? They would critically assess ethical dilemmas and value clashes working on the three scales of time and analytical level of abstraction that I have considered in this study (the micro, meso and macro) inside organisations and businesses, outside as independent auditors, and with policymakers when policy and law are negotiated. In terms of the power structure for the independent governance of these, states have stakes in this that are too high, and certainly the industry
never asked for or intended to administer such a function. Therefore, why not give civil society organisations the chance now? Specifically, civil society and nongovernmental organisations that have proven long-term independency and dedication to the human interest: provide these civil society agents, the representatives of love and human agency, with a real governance role beyond the 'activist' role. Give them the resources to compete, to professionalise, to be the 'ethical governors' of the age of Big Data. Make the structural changes.

***


Barthes, R. (1972). Mythologies, selected and translated by A. Lavers. The Noonday Press. (Originally Published in French 1957)


https://www.fastcompany.com/90396218/google-quantum-supremacy-future-ai-humanity


Brey, A. Feenberg (Eds.), Modernity and Technology (pp. 185-225). MIT Press.


Johnson, B. (11 January 2010) Privacy is no longer a social norm, says Facebook founder, the Guardian. https://www.theguardian.com/technology/2010/jan/11/facebook-privacy


https://doi.org/10.1080/1472586X.2011.548485


https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm


https://doi.org/10.1177/026327602761899246.


Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the
public sector. The Alan Turing Institute
https://doi.org/10.5281/zenodo.3240529


https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detectivestory/?mtrref=www.google.com&gwh=DC6348FBE0A5C7D9B1A6A287C0El&gwt=pay&assetType=REGIWALL


Lynsky, D., (9th October 2019). ‘Alexa, are you invading my privacy?’ – the dark side of our voice assistants. theGuardian.


https://doi.org/10.1111/j.1749-5687.2010.00109.x


HIGH-LEVEL EXPERT GROUP ON AI DOCUMENTS


POLICY DOCUMENTS

Council of Europe (1997) Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of
https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/164

Data Ethics Expert Group (2018). Data for the Benefit of the People Recommendations from the Danish Expert Group on Data Ethics. 
https://dataetiskraad.dk/sites/default/files/2020-02/Recommendations%20from%20the%20Danish%20Expert%20Group%20on%20Data%20Ethics.pdf

https://ec.europa.eu/transport/themes/infrastructure_en

https://ec.europa.eu/transport/themes/infrastructure_en


Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)


Jeg elsker min familie. Fra clara
ACKNOWLEDGEMENTS

This work is dedicated to Clara, Francesco Lapenta, Tina Pepe, Bo Erik Hasselbalch and Gry Christensen Medonos who showed me what ethics is in practice in each their uniquely human way.

Dedicato a Clara, Francesco Lapenta, Tina Pepe, Bo Erik Hasselbalch e Gry Christensen Medonos che mi hanno mostrato ognuno a loro modo cosa sia l’etica nella vita quotidiana.

Thank you:

Supervisor of the PhD thesis, Jens-Erik Mai, for the clarity and for seeing the meaning in the anarchy.

Pernille Tranberg for the force!

To my family and extended family of dear friends.
THE ARTICLES
1. Making Sense of Data Ethics. The Powers behind the data ethics debate in European policymaking


Abstract

This article offers an analytical investigation of the different actors and forces that mould definitions of “data ethics” in European policy-making. It details how data ethics public policy initiatives took shape in the context of the European General Data Protection reform, and addresses the general uncertainty that exists regarding their role and function. The paper also presents an analytical framework for an action-oriented “data ethics of power” that aims to elucidate the power relations of the ‘Big Data Society’, arguing that we recognise data ethics policy initiatives as open-ended spaces of negotiation among different interest groups that seek to guide the cultural definition of “data ethics”, with complex power relations exercised via cultural positioning.

INTRODUCTION

January 2018: The tweet hovered over my head: “Where are the ethicists?” I was on a panel in Brussels about data ethics and this wasn’t the first time a panel or initiative as such was questioned. There wasn’t the foundation proper, the right expertise was not included - the ethicists were missing, the humanists were missing, the legal experts were missing. The results, outcome and requirements of these initiatives were unclear. Would they water down the law? I understood the critiques though. How could we talk about data ethics when a law was just passed following a lengthy negotiation process on this very topic? What was the function of these discussions? If we were not there to acknowledge a consensus, that is, the
legal solution, what then was the point?

In the slipstream of sweeping data protection law reform in Europe, discussions regarding data ethics has gained traction in European public policy-making. Numerous data ethics public policy initiatives have been created, moving beyond issues of mere compliance with data protection law to increasingly focus on the ethics of big data, especially concerning private companies’ and public institutions’ handling of personal data in digital forms. Reception in public discourse has been mixed. Although gaining significant public attention and interest, these data ethics policy initiatives have also been depicted as governmental “toothless wonders” (e.g., Hill, 24 November 2017) and a waste of resources, and have been criticised for drawing attention away from public institutions’ mishandling of citizens’ data (e.g., Ingeniøren’s managing panel, op ed, 16 March 2018) and for potential “ethics washing” (Wagner, 2018), questioning the expertise and interests involved in the initiatives, as well as their normative ethics frameworks.

This article constitutes an analytical investigation of the various dimensions and actors that shape definitions of data ethics in European policy-making. Specifically, I explore the role and function of European data ethics policy initiatives and present an argument regarding how and why they took shape in the context of a European data protection regulatory reform. The explicit use of the term “ethics” calls for a philosophical framework; the term “data” for a contemporary perspective of the critical role of information in a digitalised society; and the policy context for consensus-making and problem solving. Together, these views on the role of the data ethics policy initiatives are highly pertinent. However, taken separately they each provide a one-sided kaleidoscopic insight into their role and function. For example, a moral philosophical view concerning data ethics initiatives (in public policy-making as well as in the private industry) might not be vigilant of the embedded interests and power relations; pursuit of actionable policy results may overlook their function as spaces of negotiation and positioning; while viewing data ethics initiatives as something radically new in the age of big data can lose sight of their place in and relation to history and governance in general.
In my analysis, I therefore adopt an interdisciplinary approach that draws on methods and theories from different subfields within applied ethics, political science, sociology, culture and infrastructure/STS studies. A central thesis of this article is that we should perceive data ethics policy initiatives as open-ended spaces of negotiation embedded in complex socio-technical dynamics, which respond to multifaceted governance challenges extended over time. This is why, we should not view data ethics policy initiatives as solutions in their own right. They do not replace legal frameworks such as the European General Data Protection Regulation (GDPR). Rather, they complement existing law and may inspire, guide and even set in motion political, economic and educational processes that could foster an ethical “design” of the big data age, covering everything from the introduction of new laws, the implementation of policies and practices in organisations and companies and the development of new engineering standards, to awareness campaigns among citizens and educational initiatives.

In the following, I first outline a cross-disciplinary conceptualisation of data ethics, presenting what I define as an analytical framework for a data ethics of power. I then describe the data ethics public policy focus in the context of the GDPR. I recognise that ethics discussions are implicit in legislative processes. Nevertheless, in this article I do not specifically focus on the regulation’s negotiation process as such, but rather on policymakers’ explicit use of the term “data ethics”, and especially on the emergence of formal data ethics policy initiatives (for instance, committees, working groups, stated objectives and results), many of which followed the adoption of the GDPR. I subsequently move on to an analysis of data ethics as described in public policy reports, statements, interviews and events in the period 2015–2018. In conclusion, I take a step back and review the definition of data ethics. Today, data ethics is an idea, concept and method that is used in policy-making, but which has no shared definition. While more aligned conceptualisations of data ethics might provide a guiding step for a collective vision for actions in law, business and society in general, an argument that runs through this article is that there is no definition of data ethics in this space neutral of values and politics. Therefore, we must position ourselves within a context-specific type of ethical action.
This article is informed by a study that I am conducting on data ethics in governance and technology development in the period 2017-2020. In that study and this article, I use an ethnographically informed approach based on active and embedded participation in various data protection/internet governance policy events, working groups and initiatives. Qualitative embedded research entails an immersion of the researcher in the field of study as an active and engaged member to achieve thorough knowledge and understanding (Bourdieu, 1997; Bourdieu & Wacquant 1992; Goffman, 1974; Ingold, 2000; Wong, 2009). Thus, essential to my understanding of the underlying dimensions of the topic of this article is my active participation in the internet governance policy community. I was for example part of the Danish government’s data ethics expert committee (2018) and am part of the European Commission’s Artificial Intelligence High Level Expert group (2018-2020). I am also the founder of the non profit organisation DataEthics.eu, which is active in the field.

In this article, I also draw on ideas, concepts and opinions generated in interaction with nine active players (decision-makers, policy advisors and civil servants) whom contributed to my understanding of the policy-making dynamics by sharing their experiences with data ethics in European policy-making (see further in references). The interviewees were informed about the study and that they would not be represented by name and institution in any publications, as I wanted them to be minimally influenced by institutional interests and requirements in their accounts.

WHAT IS DATA ETHICS? A DATA ETHICS OF POWER

In this section I introduce the emerging field of data ethics as the cross-disciplinary study of the distribution of societal powers in the socio-technical systems that form the fabric of the “Big Data Society”. Based on theories, practices and methods within applied ethics, legal studies and cultural studies, social and political sciences, as well as a movement within policy and business, I present an analytical framework for a “data ethics of power”.

As a point of departure, I define a data ethics of power as an action-oriented analytical framework concerned with making visible the power
relations embedded in the “Big Data Society” and the conditions of their
negotiation and distribution, in order to point to design, business, policy,
social and cultural processes that support a human-centric distribution of
power. In a previous book (Hasselbalch and Tranberg, 2016) we described
data ethics as a social movement of change and action: “Across the globe,
we’re seeing a data ethics paradigm shift take the shape of a social
movement, a cultural shift and a technological and legal development that
increasingly places the human at the centre” (p. 10). Thus, data ethics can
be viewed as a proactive agenda concerned with shifting societal power
relations and with the aim to balance the powers embedded in the Big Data
Society. This shift is evidenced in legal developments (such as the GDPR
negotiation process) and in new citizen privacy concerns and practices such
as the rise in use of ad blockers and privacy enhancing services, etc. In
particular, new types of businesses emerge that go beyond mere
compliance with data protection legislation when incorporating data
ethical values in collection and processing of data, as well as their general
innovation practices, technology development, branding and business
policies.

Here, I use the notion of “Big Data Society” to reflectively position data
ethics in the context of a recent data (re)evolution of the “Information
Society”, enabled by computer technologies and dictated by a
transformation of all things (and people) into data formats (“datafication”)
in order to “quantify the world” (Mayer-Schonberger & Cukier, 2013, p. 79)
to organise society and predict risks. While I suggest that this is not an
arbitrary evolution, but can also be viewed as an expression of negotiations
between different ontological views on the status of the human being and
the role of science and technology. As the realisation of a prevailing
ideology of modernist scientific practices to command nature and living
things, the critical infrastructures of the Big Data Society may therefore
very well be described as modernity embodied in a “lived reality” (Edwards,
2002, p. 191) of control and order. From this viewpoint, a data ethics of
power can be described as a type of post-modernist, or in essence vitalist,
call for a specific kind of “ethical action” (Frohmann, 2007, p. 63) to free
the living/human being from the constraints of the practices of control
embedded in the technological infrastructures of modernity that at the
same time reduce the value of the human being. It is here valuable to
understand current calls for data ethical action in extension of the philosopher Henri Bergson’s vitalist arguments at the turn of the last century against the scientific rational intellect that provides no room for, or special status to, the living (Bergson, 1988, 1998). In a similar ethical framework, Gilles Deleuze, who was also greatly inspired by Bergson (Deleuze, 1988), later described over-coded “Societies of Control” (Deleuze, 1992), which reduce people (“dividuals”) to a code marking their access and locking their bodies in specific positions (p. 5). More recently, Spiekerman et al. (2017) in their Anti-Transhumanist Manifesto directly oppose a vision of the human as merely information objects, no different than other information objects (that is; non-human informational things), which they describe as “an expression of the desire to control through calculation. Their approach is limited to reducing the world to data-based patterns suited for mechanical manipulation” (p. 2).

However, a data ethics of power should also be viewed as a direct response to the power dynamics embedded in and distributed via our very present and immediate experiences of a “Liquid Surveillance Society” (Lyon, 2010). Surveillance studies scholar David Lyon (2014) envisions an “ethics of Big Data practices” (2014, p. 10) to renegotiate what is increasingly exposed to be an unequal distribution of power in the technological big data infrastructures (add “co-creation”, Lyon, 2018). Within this framework we do not only pay conventional attention to the state as the primary power actor (of surveillance), but also include new stakeholders that gain power through accumulation and access to big data. For example, in the analytical framework of a data ethics of power, changing power dynamics are progressively more addressed in the light of the information asymmetry between individuals and the big data companies that collect and process data in digital networks (Pasquale, 2015; Powles, 2015–2018; Zuboff, 5 March 2016, 9 September 2014, 2019).

Beyond this fundamental theoretical framing, a data ethics of power can be explored in an interdisciplinary field addressing the distribution of power in the Big Data Society in diverse ways.

For instance, in a computer ethics perspective, power distributions are approached as ethical dilemmas or as implications of the very design and practical application of computer technologies. Indeed, technologies are
never neutral, they embody moral values and norms (Flanagan, Howe, & Nissenbaum, 2008), hence power relations can be identified through analysing how technologies are designed in ethical or ethically problematic ways. Information science scholars Batya Friedman and Helen Nissenbaum (1996) have illustrated different types of bias embedded in existing computer systems that are used for tasks such as flight reservations and the assignment of medical graduates to their first job, and have presented a framework for such issues in the design of computer systems. From this perspective, we can also describe data ethics as what the philosophy and technology scholar Philip Brey terms a “Disclosive Computer Ethics”, identifying moral issues such as “privacy, democracy, distributive justice, and autonomy” (Brey, 2000, p. 12) in opaque information technologies. Phrased differently, a data ethics of power presupposes that technology has “politics” or embedded “arrangements of power and authority” (Winner, 1980, p. 123). Case studies of specific data processing software and their use can be defined as data ethics case studies of power, notably the “Machine Bias” study (Angwin et al., 2016), which exposed discrimination embedded in data processing software used in United States defence systems, and Cathy O’Neil’s (2016) analysis of the social implications of the math behind big data decision-making in everything from getting insurance, credit to getting and holding a job.

Nevertheless, data systems are increasingly ingrained in society in multiple forms (from apps to robotics) and have limitless and wide-ranging ethical implications (from price differentiation to social scoring), necessitating that we look beyond design and computer technology as such. Data ethics as a recent designation represents what philosophers Luciano Floridi and Mariateresa Taddeo (2016, p. 3) describe as a primarily semantic shift within a computer and information ethics philosophical tradition from a concern with the ethical implications of the “hardware” to one with data and data science practices. However, looking beyond applied ethics in the field of philosophy to a data ethics of power, our theorisation of the Big Data Society is more than just semantic. The conceptualisation of a data ethics of power can also be explored in a legal framework, as an aspect of the rule of law and protection of citizens’ rights in an evolving Big Data Society. Here, redefining the concept of privacy (Cohen, 2013; Solove, 2008) in a legal studies framework, addresses the ethical implications of
new data practices and configurations that challenge existing laws, and thereby the balancing of powers in a democratic society. As legal scholars Neil M. Richards and Jonathan King (2014) argue: “Existing privacy protections focused on managing personally identifying information are not enough when secondary uses of big data sets can reverse engineer past, present, and even future breaches of privacy, confidentiality, and identity” (p. 393). Importantly, these authors define big data “socially, rather than technically, in terms of the broader societal impact they will have,” (Richards & King, 2014, p. 394) providing a more inclusive analysis of a “big data ethics” (p. 393) and thus pointing to the ethical implications of the empowerment of institutions that possess big data capabilities at the expense of “individual identity” (p. 395).

Looking to the policy, business and technology field, the ethical implications of the power of data and data technologies are framed as an issue of growing data asymmetry between big data institutions and citizens in the very design of data technologies. For example, the conceptual framework of the “Personal Data Store Movement” (Hasselbalch & Tranberg, 27 September 2016) is described by the non-profit association MyData Global Movement as one in which “[i]ndividuals are empowered actors, not passive targets, in the management of their personal lives both online and offline – they have the right and practical means to manage their data and privacy” (Poikola, Kuikkanemi, & Honko, 2018). In this evolving business and technology field, the emphasis is on moving beyond mere legal data protection compliance, implementing values and ethical principles such as transparency, accountability and privacy by design (Hasselbalch & Tranberg, 2016), and ethical implications are mitigated by values-based approaches to the design of technology. For example, engineering standards such as those of IEEE P7000s Ethics and AI standards 3 that seek to develop ethics by design standards and guiding principles for the development of artificial intelligence (AI). A values based design approach is also revisited in recent policy documents such as section 5.2. “Embedded values in technology – ethical-by-design” of the European Parliament’s “Resolution on Artificial Intelligence and Robotics” adopted in February 2019.

A key framework for data ethics is the ‘human-centric approach’ that we
increasingly see included within ethics guidelines and policy documents. For example, the European Parliament’s (2019, V.) resolution states that “whereas AI and robotics should be developed and deployed in a human-centred approach with the aim of supporting humans at work and at home...”. The EC High Level Expert Group on Artificial Intelligence’s draft ethics guidelines also stress how the ‘human-centric approach’ to AI is one that “strives to ensure that human values are always the primary consideration” (working document, 18 December 2018, p. iv), and directly associate it with the balance of power in democratic societies: “political power is human centric and bounded. AI systems must not interfere with democratic processes” (p. 7). The ‘human-centric approach’ in European policy-making is framed in a European fundamental rights framework (as for example extensively described in the European Commission’s AI High Level Expert group’s draft ethics guidelines) and/or with an emphasis on the human being’s interests prevailing over “the sole interests of society or science” (article 2, “Oviedo Convention”). Practical examples of the ‘human-centric approach’ can also be found in technology and business developments that aim to preserve the specific qualities of humans in the development of information processing technologies. Examples include the Human in the Loop (HITL) approach to the design of AI, The International Organization for Standardization (ISO) standards on human-centred design (HCD) and the Personal Data Store Movement, which is defined as “A Nordic Model for human-centered personal data management and processing.” (Poikola et al., 2018)

EUROPEAN DATA ETHICS POLICY INITIATIVES IN CONTEXT
Policy debates that specifically address ethics in the context of technological developments have been ongoing in Europe since the 1990s. The debate has increasingly sought to harmonise national laws and approaches in order to preserve a European value framework in the context of rapid technological progress. For instance, the Council of Europe’s “Oviedo Convention” was motivated by what Wachter (1997, p. 14) describes as “[t]he feeling that the traditional values of Europe were threatened by rapid and revolutionary developments in biology and medicine”. Data ethics per se gained momentum in pan-European politics in the final years of the negotiation of the GDPR, through the establishment of a number of initiatives directly referring to data and/or digital ethics.
Thus, the European Data Protection Supervisor (EDPS) Digital Ethics Advisory Group (2018, p. 5) describes its work as being carried out against “a growing interest in ethical issues, both in the public and in the private spheres and the imminent entry into force of the General Data Protection Regulation (GDPR) in May 2018”.

Examination of the differences in scope and the stakeholders involved in respectively the development of the 1995 Data Protection Directive and the negotiation process of the GDPR beginning with the European Commission’s proposal in 2012, provides some insight into the evolution of the focus of data ethics. The 1995 Directive was developed by a European working party of privacy experts and national data protection commissioners in a process that excluded business stakeholders (Heisenberg, 2005). Nevertheless, the group of actors influencing and participating in the development of the GDPR process progressively expanded, with new stakeholders comprising consumer and civil liberty organisations and American industry representatives and policymakers. The GDPR was generally described as one of the most lobbied EU regulations (Warman, 8 February 2012). At the same time, the public increasingly scrutinised the ethical implications of a big data era, with numerous news stories published on data leaks and hacks, algorithmic discrimination and data-based voter manipulation.

Several specific provisions of the GDPR were discussed inside and outside the walls of European institutions. For example, the “right to erasure” proposed in 2012 was heavily debated by industry and civil society organisations, especially in Europe and the USA, and was frequently described in the media as a value choice between privacy and freedom of expression. In 2013, the transfer of data to third countries (including those covered by the EU-US Safe Harbour agreement) engendered a wider public debate between certain EU parliamentarians and US politicians regarding mass surveillance and the role of large US technology companies. Another example was the discussion of an age limit of 16. This called civil society advocates into action (Carr, Should I laugh, cry or emigrate?, 13 December 2015) and led to new alliances with US technology companies regarding young people’s right to “educational and social opportunities” (Richardson, “European General Data Protection Regulation draft: the debate”, 10...
December 2015). A last-minute decision rendered it possible to lower the age limit to 13 in member states.

These intertwined debates and negotiations illustrate how the data protection field was transformed within a global information technology infrastructure. It took shape as a negotiation of competing interests and values between economic entities, EU institutions, civil society organisations, businesses and third country national interests. We can also perceive these spaces of negotiation of rights, values and responsibilities and the creation of new alliances to have a causal link with the emergence of data ethics policy initiatives in European policy-making. In the years following the first communication of the reform, data protection debates were extended, with the concept of data ethics increasingly included in meeting agendas, debates in public policy settings and reports and guidelines. Following the adoption of the GDPR, the list of European member states or institutions with established data or digital ethics initiatives and objectives rapidly grew. Examples included the UK government’s announcement of a £9 million Centre for Data Ethics and Innovation with the stated aim to “advise government and regulators on the implications of new data-driven technologies, including AI” (Digital Charter, 2018). The Danish government appointed a data ethics expert committee 4 in March 2018 with a direct economic incentive to create data ethics recommendations to Danish industry and to turn responsible data sharing into a competitive advantage for the country (Danish Business Authority, 12 March 2018). Several member states’ existing and newly established expert and advisory groups and committees began to include ethics objectives into their work. For example, the Italian government established an AI Task Force in April 2017, publishing its first white paper in 2018 (AI Task Force/Italy, 2018) with an explicit section on ethics. The European Commission’s communication on an AI strategy, published in April 2018, also included the establishment of an AI High Level Expert Group 5, whose responsibility it was, among others, to publish ethics guidelines for AI in Europe the following year.

DATA ETHICS - POLICY VACUUMS

“I’m pretty convinced that the ethical dimension of data protection and privacy protection is going to become a lot more important in the years to
come” (in ‘t Veld, 2017). These words of a European parliamentarian in a public debate in 2017 referred to the evolution of policy debates regarding data protection and privacy. You can discuss legal data protection provisions, she claimed, but then there is “a kind of narrow grey area where you have to make an ethical consideration and you say what is more important” (in ‘t Veld, 2017). What did she mean by her use of the term “ethics” in this context?

In an essay entitled “What is computer ethics?” (1985), the moral philosophy scholar James H. Moor described the branch of applied ethics that studies the ethical implications of computer technologies. Published only a few years after Acorn, the first IBM personal computer, was introduced to the mass market, Moor was interested in computer technologies per se (what is special about computers), as well as the policies required in specific situations where computers alter the state of affairs and create something new. But he also predicted a more general societal revolution (Moor, 1985, p. 268) due to the introduction of computers that will “leave us with policy and conceptual vacuums” (p. 272). Policy vacuums, he argued, would present core problems and challenges, revealing “conceptual muddles” (p. 266), uncertainties and the emergence of new values and alternative policies (p. 267).

If we view data ethics policy initiatives according to Moor’s framework, they can be described as moments of sense-making and negotiation created in response to the policy vacuums that arise when situations and settings are amended by computerised systems. In an interview conducted at the Internet Governance Forum (IGF) in 2017, a Dutch parliamentarian described how in 2013, policy-makers in her country rushed to tackle the transformations instigated by digital technologies that were going “very wrong” (Interview, IGF 2017). In response, she proposed the establishment of a national commission to consider the ethical challenges of the digital society: “it’s very hard to get the debate out of the trenches, you know, so that people stop saying, ‘well this is my position and this is my position’, but to just sit back and look at what is happening at the moment, which is going to be so huge, so incredible, we have no idea what is going to happen with our society and we need people to think about what to do about all of this, not in the sense you know, ‘I don’t want it’, but more in the sense, ‘are
there boundaries?’ ‘Do we have to set limits to all of these possibilities that will occur in the coming years?’” Similarly, in another interview conducted at the same event, a representative of a European country involved in the information policy of the Committee of Ministers of the Council of Europe discussed how the results of the evolution of the Information Society included “violations”, “abuses” and recognition of the internet’s influence on the economy. Concluding, she stated that: “We need to slow down a little bit and to think about where we are going”.

In reviewing descriptions of data ethics initiatives, we can note implicit acknowledgement of the limits of data protection law in harnessing all of the ethical implications of a rapidly evolving information and data infrastructure. Data ethics thus become a means to make sense of emerging problems and challenges and to evaluate various policies and solutions. For example, a report from EDPS from 2015 states: “In today’s digital environment, adherence to the law is not enough; we have to consider the ethical dimension of data processing” (p. 4). It continues by describing how different EU law principles (such as data minimisation and the concepts of sensitive personal data and consent) are challenged by big data business models and methods.

The policy vacuums described in such reports and statements highlight the uncertainties and questions that exist regarding the governance of a socio-technical information infrastructure that increasingly shapes not only personal, but also social, cultural and economic activities.

In the same year as Moor’s essay was published, communications scholar Joshua Meyrowitz’s No Sense of Place (1985) portrayed the emergence of “information systems” that modify our physical settings via new types of access to information, thereby restructuring our social relations by transforming situations. As Meyrowitz (1985, p. 37) argued, “[w]e need to look at the larger, more inclusive notion of “patterns of information””, illustrating how our information realities have real qualities that shape our social and physical realities. Accordingly, European policymakers emphasise the real qualities of information and data. They see digital data processes as meaningful components of social power dynamics. Information society policy-making thus becomes an issue of the distribution of resources and of social and economic power, as an EU
Competition Commissioner stated at a DataEthics.eu event on data as power in Copenhagen in 2016: “I’m very glad to have the chance to talk with you about how we can deal with the power that data can give” (Vestager, 9 September 2016). Thus, data ethics policy debates have moved beyond the negotiation of a legal data protection framework, increasingly involving a general focus on information society policy-making, in which different sectional policy areas are intertwined. As the European Commissioner for Competition elaborated at the DataEthics.eu event: “So competition is important. It keeps the pressure on companies to give people what they want. And that includes security and privacy. But we can’t expect competition enforcement to solve all our privacy problems. Our first line of defence will always be rules that are designed specifically to guarantee our privacy”.

DATA ETHICS - CULTURE AND VALUES

According to Moor, the policy vacuums that emerge when existing policies clash with technological evolution, force us to “discover and make explicit what our value preferences are” (1985, p. 267). He proposes that the computer induced societal revolution will occur in two stages, marked by the questions that we ask. In the first “Introduction Stage”, we ask functional questions: How well does this and that technology function for its purpose? In the second “Permeation Stage”, when institutions and activities are transformed, Moor argues that we will begin to ask questions regarding the nature and value of things (p. 271). Such second-stage questions are echoed in the European policy debate of 2017, as one Member of the European Parliament (MEP) who was heavily involved in the GDPR negotiation process argued in a public debate: “[this is] not any more a technical issue, it’s a real life long important learning experience” (Albrecht, 2017), or as another MEP claimed in the same debate: “The GDPR is not only a legislative piece, it’s like a textbook, which is teaching us how to understand ourselves in this data world and how to understand what are the responsibilities of others and what are the rules which is governing in this world” (Lauristin, 2017).

Consequently, the technicalities of new data protection legislation are transformed into a general discussion about the norms and values of a big data age. Philip Brey describes values as “idealized qualities or conditions
in the world that people find good”, ideals that we can work towards realising (2010, p. 46). However, values are not just personal ideals; they are also culturally situated. The cultural theorist Raymond Williams (1958, p. 6) famously defined culture as a “shape”, a set of purposes and common meanings expressed “in institutions, and in arts and learning”, which emerge in a social space of “active debate and amendment under the pressures of experience, contact and discovery”. Culture is thus traditional as well as creative, consisting of prescribed dominant meanings and their negotiation (Williams, 1958). Similarly, the anthropologist James Clifford (1997) replaced the metaphor of “roots” (an image of the original, authentic and fixed cultural entity) with “routes”: intervals of negotiation and translation between the fixed cultural points of accepted meaning. Values are advanced in groups with shared interests and culture but they exist in spaces of constant negotiation. In an interview conducted at the IGF 2017, one policy advisor to an MEP enquired as to the role of values in the GDPR’s negotiations, described privacy as a value shared by a group of individuals involved in the reform process: “I think a group of core players shared that value (…) all the way from people who wrote the proposal at the Commission, to the Commissioner in charge to the rapporteur from the EU Parliament, they all (…) to some extent shared this value, and I think that they managed to create a compromise closer to their value than to others”. He also explained how discussions about values were emerging in processes of negotiation between diverse and at times contradictory interests: “the moment you see a conflict of interest, that is when you start looking at the values (…) normally it would be a discussion about different values (…) an assessment of how much one value should go before another value (…) so some people might say that freedom of information might be a bigger value or the right to privacy might be a bigger value”.

Accordingly, ethics in practice, or what Brey refers to as “the act of valuing something, or finding it valuable (…) to find it good in some way” (2010, p. 46) is in essence never merely a subjective practice, but neither is it a purely objective construct. If we investigate the meaning of data ethics and ethical action in European data protection policy-making, we can see the points of negotiation. That is, if we look at what happens in the “intervals” between established value systems and the renegotiation of these in new contexts, we discover clashes of values and negotiation as well
as the contours of cultural positioning.

**DATA ETHICS - POWER AND POSITIONING**

Philosophy and media studies scholar Charles Ess (2014) has illustrated how culture plays a central role in shaping our ethical thinking about digital technologies. For instance, he argues that people in Western societies place ethical emphasis on “the individual as the primary agent of ethical reflection and action, especially as reinforced by Western notions of individual rights” (p. 196). Such cultural positioning in a global landscape can also be identified in the European data ethics policy debate. An example is the way in which one participant in the 2017 MEP debate discussed above described the GDPR with reference to the direct lived experiences of specific European historical events: “It is all about human dignity and privacy. It is all about the conception of personality which is really embedded in our culture, the European culture (...) it came from the general declaration of human rights. But there is a very very tragic history behind war, fascism, communism and totalitarian societies and that is a lesson we have learned in order to understand why privacy is important” (Lauristin, 2017).

Values such as human dignity and privacy are formally recognised in frameworks of European fundamental rights and data protection law, and conscious of their institutionalised roots in the European legal framework, European decision-makers will reference them when asked about the values of “their” data ethics. Awareness of data ethics thus becomes a cultural endeavour, transferring European cultural values into technological development. As stated in an EDPS report from 2015: “The EU in particular now has a ‘critical window’ before mass adoption of these technologies to build the values into digital structures which will define our society” (p. 13).

When exploring European data ethics policy initiatives as spaces of value negotiations, a specific cultural arrangement emerges. In this context, policy and decision-makers position themselves against a perceived threat to a specifically European set of values and ethics that is pervasive, opaque and embedded in technology. In particular, a concern with a new opponent to the state power emerges. In an interview conducted in 2018 at an institution in Europe, a project officer reflected on her previous work in a
European country’s parliament and government where concerns with the alternative form of power that the internet represents had surfaced. The internet is the place where discussions are held and decisions are made, she said, before remembering the policy debates concerning “GAFA” (the acronym for the four giant technology companies of Google, Apple, Facebook and Amazon). Such a clash in values has been directly addressed by European policymakers in public speeches and debates, increasingly naming the technology industry stakeholders they deem responsible. Embedded values of technology innovation are a “wrecking ball”, aiming not simply to “play with the way society is organised but instead to demolish the existing order and build something new in its place”, argued a President of the European Parliament in a speech in 2016 (Schultz, 2016).

Values and ethics are hence directly connected with a type of cultural power that is built into technological systems. As one Director for Fundamental Rights and Union Citizenship, European Commission DG Justice claimed in a 2017 public debate: “the challenge of ethics is not in the first place with the individual, the data subject; the challenge is with the controllers, which have power, they have power over people, they have power over data, and what are their ethics? What are the ethics they instil in their staff? In house compliance ethics? Ethics of engineers?” (Nemitz, 2017).

DATA ETHICS - SPACES OF NEGOTIATION

When dealing with the development of technical systems, we are inclined towards points of closure and stabilisation (Bijker et al., 1987) that will guide the governance, control and risk mitigation of the systems. Relatedly, we can understand data ethics policy initiatives as end results with the objectives “to formulate and support morally good solutions (e.g., right conducts or right values)” (Floridi & Taddeo, 2016, p. 1), emphasising algorithms (or technologies) that may not be “ethically neutral” (Mittelstadt et al., 2016, p. 4). That is to say, as solutions to the ethical problems raised within the very design of technologies, the data processing activities of the algorithms or the collection and dissemination of data. However, I would like to address data ethics policy initiatives in their contexts of interest and value negotiation. For instance, where does morality begin and end in a socio-technical infrastructure that extends across jurisdictions and continents, cultural value systems and societal sectors?
The technical does indeed in the very design represent forms of order, as the political theorist Langdon Winner reminded us (1980, p. 123). That is, it is “political” and thus has ethical implications when creating by design “wonderful breakthroughs by some social interests and crushing setbacks by others” (Winner, 1980, p. 125). To provide an example, the Facebook APIs that facilitated the mass collection of user data, before these were reused and processed by Cambridge Analytica, were specifically designed to track users and share data en masse with third parties, hence directly enabling the mass collection, storage and processing of data. However, these design issues of the technical are also “inextricably bound up into an organic whole” with economic, social, political and cultural problems (Callon, 1987, p. 84). An analysis of data ethics as it is evolving in the European policy sphere demonstrates the complexity of governance challenges arising from the infrastructure of the information age being “shaped by multiple agents with competing interests and capacities, engaged in an indefinite set of distributed interactions over extended periods of time” (Harvey et al., 2017, p. 26). Governance in this era is, as highlighted by internet governance scholars Jeanette Hofmann et al., a “heterogeneous process of ordering without a clear beginning or endpoint” (2016, p. 1412). It consists of actors engaged in “fields of struggle” (Pohle et al., 2016) of meaning making and competing interpretations of policy issues that are “continuously produced, negotiated and reshaped by the interaction between the field and its actors” (p. 4). I propose that we also explore, as essential components of our data ethics endeavours, the complex dynamics of the ways in which powers are distributed and how interests are met in spaces of negotiation.

Evidently, we must also recognise data ethics policy initiatives as components of a general infrastructural development’s rhythm rather than caved in ethical solutions and isolated events. Understand them as the kind of negotiation posts that repeatedly occur throughout the course of a technological system’s development (Bijker et al., 1987), and as segments of a process of standardisation and consensus-building within a complex general technological evolution of our societies that “contain messy, complex, problem-solving components” (Hughes, 1987, p. 51). The technological systems of modernity are like the architecture of mundane buildings. They reside, as Edwards (2002, p. 185) claims, in a “naturalised
background, ordinary as trees, daylight, and dirt”. Silently they represent, constitute and are constituted by both our material and imagined modern societies and the distribution of power within. They remain unnoticed until they fail (Edwards, 2002). But when they do fail, we see them in all their complexity. An example is the US intelligence officers PowerPoint presentations (The Guardian, 2013) detailing the “PRISM program” leaked by Edward Snowden in 2013 that provide a detailed map of an information and data infrastructure that is characterised by intricate interconnections between a state organisation of mass surveillance, laws, jurisdictions and budgets, and the technical design of the world wide web and social media platforms. The technological infrastructures are indeed like communal buildings. With doors that we never give a second thought until the day we find one of them locked.

**CONCLUSION**

October 2018: “These are just tools!” one person exclaimed. We were at a working group meeting where an issue with using Google Docs for the practical work of the group was raised and discussed at length. While some were arguing for an official position on the use of the online service, mainly with reference to what they described as Google’s insufficient compliance with European data protection law, others saw the discussion as a waste of time. Why spend valuable work time on this issue?

What is data ethics? Currently, the reply is shrill, formally framed in countless statements, documents and mission statements from a multitude of sources, including governments, intergovernmental organisations, consultancy firms, companies, non-governmental organisations, independent experts and academics. But it also emerges when least expected, in “non-allocated” moments of discussion. Information technologies that permeate every aspect of our lives today, from micro work settings to macro economics and politics, are increasingly discussed as “ethical problems” (Introna, 2005, p. 76) that must be solved. Their pervasiveness sparks moments of ethical thinking, negotiated in terms of moral principles, values and ideal conditions (Brey, 2010). In allocated or unallocated spaces of negotiation, moments of pause and sense-making (Moor, 1985), we discuss the values (Flanagan et al., 2008) and politics (Winner, 1980) of the business practices, cultures and legal jurisdictions that
shape them. These spaces of negotiation encompass very concrete discussions regarding specific information technology tools, but increasingly they also evolve into reflections concerning general challenges to established legal frameworks, individuals’ agency and human rights, as well as questions regarding the general evolution of society. As one Danish minister said at the launch of a national data ethics expert group: “This is about what society we want” (Holst, 11 March 2018).

In this article, I have explored data ethics in the context of a European data protection legal reform. In particular, I have sought to answer the question: “What is data ethics?” with the assumption that the answer will shape how we perceive the role and function of data ethics policy initiatives. Based on a review of policy documents, reports and press material, alongside analysis of the ways in which policymakers and civil servants make sense of data ethics, I propose that we recognise these initiatives as open-ended spaces of negotiation and cultural positioning.

This approach to ethics might be criticised as futile in the context of policy and action. However, I propose that understanding data ethics policy initiatives as spaces of negotiation does not prevent action. Rather, it forces us to make apparent our point of departure: the social and cultural values and interests that shape our ethical action. We can thus create the potential for a more transparent negotiation of ethical action in the “Big Data Era”, enabling us to acknowledge the macro-level data ethics spaces of negotiation that are currently emerging not only in Europe but globally.

This article’s analytical investigation of European data ethics policy initiatives as spaces of cultural value negotiations has revealed a set of actionable thematic areas. It has illustrated a clash of values and an emerging concern with the alternative forms of power and control embedded in our technological environment, which exert pressure on people and individuals in particular. Here, a data ethics of power that takes its point of departure in Gilles Deleuze's description of computerised Societies of Control (1992) enables us to think about the ethical action that is necessary today. Ethical action could for example concern the empowerment of individuals to challenge the laws and norms of opaque algorithmic computer networks, as we have noted in debates on the right
to explanation and the accountability and interpretability of algorithms. Ethical action may also strive towards ideals of freedom in order to break away from coding, to become indiscernible to “Weapons of Math Destruction” (O’Neil, 2016) that increasingly define, shape and limit us as individuals, as seen for instance in the digital self-defence movement (Heuer & Tranberg, 2013). Data ethics missions such as these are rooted in deeply personal experiences of living in coded networks, but they are also based on growing social and political movements and sentiments (Hasselbalch & Tranberg, 2016).

Much remains to be explored and developed regarding the power dynamics embedded in the evolving data ethics debate, not only in policy-making, but also in business, technology and public discourse in general. This article seeks to open up a more inclusive and holistic discussion of data ethics in order to advance investigation and understanding of the ways in which values are negotiated, rights and authority are distributed, and conflicts are resolved.

References


Floridi, L., & Taddeo, M. (2016). What is data ethics?. Philosophical


**POLICY DOCUMENTS AND REPORTS**


Footnotes

1. By “European” I am not only focusing on the European Union (EU), but on a constantly negotiated cultural context, and thus for example I do not exclude organisations like the Council of Europe or instruments such as the European Convention of Human Rights.

2. Interviews informing the article (anonymous, all audio recorded, except from one based on written notes, four directly quoted in the article): two policy advisors; four European institution officers; one data protection commissioner; one representative of a European country to the Committee of Ministers of the Council of Europe; one European parliamentarian.

3. I am the vice chair of the IEEE P7006 standard on personal data AI agents.
4. I was one of the 12 appointed members of this committee (2018).
5. I was one of the 52 appointed members of this group (2018-2020).
2. Culture by design: A data interest analysis of the European AI policy agenda

Open Access: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Abstract
This article investigates a moment of the big data age in which artificial intelligence became a fixed point of global negotiations between different interests in data. In particular, it traces and explicates cultural positioning as an interest in the artificial intelligence momentum with an investigation of the unfolding of a European AI policy agenda on trustworthy AI in the period 2018–2019.

Introduction
At the end of the first 20 years of the twenty-first century, artificial intelligence technologies (AI) [1] came to be at the center of a global public debate on policy, media and industry. From transpiring as a scientific endeavor and sci-fi curiosity, AI had transformed into socio-technical systems with rapid and broad societal adoption and consequently a fixed point of governance in the European Union. EU legislators had just implemented a momentous data protection law reform to address challenges of a big data digitalization of societies, and on a global scale, states and companies alike were carving their space with more or less aggressive data harvesting advances, while citizens were toiling to understand their own role in emerging big data technological environments. Against that background, a European AI strategy was published in early 2018 by the European Commission and further developed in policy and expert group initiatives over a two-year period with a growing emphasis on “ethical technologies” and “trustworthy AI”.
This article traces and explicates “culture” as an interest in a societal AI
momentum with an analysis of the European AI policy agenda as it evolved in the period 2018–2019, focusing in particular on the work of a high-level expert group on AI set up by the European Commission to inform the AI strategy. The article’s analysis focuses on events, documents and statements that have contributed to the development of an official AI agenda in Europe and is informed by the author’s active participation as a member of the high-level group. Predominantly, the European AI agenda is examined as a component of a general process of value negotiations in a global environment. Indeed, the evolving agenda was from the outset explicitly framed as a European “third way” in what was dubbed in public discourse the “global AI race” between Europe, and the U.S. and China.

The term AI was used in public policy-making and discourse in the 2010s generally to describe the next frontier in Big Data Society. AI was developed, designed and used by all types of societal stakeholders to make sense of large amounts of data, predict patterns, analyze risks and act on that knowledge to make decisions within politics, culture, industries, and on life trajectories. In essence, the popular use of the term came to denote a particular advanced complex design of big data systems, automated, goal-oriented, perceptive, reasoning and made powerful by complex data acquisition and processing. Thus, above all, the article investigates an institutionally framed cultural positioning as an interest in data understanding AI as complex data processing systems and data design that forms the locus of societal power dynamics. As such, it does not seek to predict the path of AI adoption as this will be shaped by a much broader sum of actors, interests and conditions, which includes the formally mitigated consequences of law, policy and institutional practice as well as the unintended outcomes of people’s (users, engineers, etc.) practices (Epstein, et al., 2016).

Theoretically, the article is grounded in a discussion of the role of culture and interests in the development and governance of socio-technical systems. It builds on conceptualizations of culture, power and technologies in cultural studies, applied ethics and science and technology studies (STS). In combination these perspectives treat technologies as dynamic concepts constantly in negotiation with human, societal and cultural factors. The understanding is that while technological artefacts may impose on humans and human societies, humans simultaneously impose on technology, and we may choose to do so with intention and direction. We create laws,
policies and standards; we educate and program, hack and revolt. This is an important view on technological development and change as it empowers human governance efforts when considering the multiple human and non-human factors that shape the direction of a technological development.

The European AI agenda: Sculpting the cultural interest in AI
Since the 1980s, AI’s adoption in society has progressed from the rule-based expert systems encoded with the knowledge of human experts to systems that evolve and learn from big data in digital environments with increasingly autonomous decision-making agency and capabilities (Alpaydin, 2016). In the 2010s, socio-technical data infrastructures enhanced by AI software systems to autonomously, or semi-autonomously, perceive and interpret their environments were worldwide increasingly embedded in private and public sectors in health care, security, finance, emergency, defence, e-government, law, transportation and energy. The U.S. had been a first mover in terms of global capital investment in AI as well as in the development of an AI ecosystem, and China rapidly followed suit (Merz, 2019). In Europe, an increasing number of examples of socially challenging applications of AI from these regions had been in the public limelight, for example, the use of a biased sentencing software in the U.S. judicial system (Angwin, et al., 2016) or the mass citizen social credit scoring system of China (Kobie, 2018). But gradually the social implications of AI used in European settings were also edging into public awareness as a component of decision-making in many different sectors [2]. The European Union was, for example, proposing and adopting initiatives to establish smart border management systems and to integrate instruments for data processing and decision-making systems in asylum and immigration and law enforcement cooperation. In Europe there were also examples of experimenting with frameworks for automating detection and analysis of terrorist-related online contents and financing activities. At the same time, individual member states were toying with AI for predictive policing, public administration of benefits, tracing vulnerable children, for tax collection purposes and even for social scoring, while private sector examples included most profoundly AI in banking and insurance (Spielkamp, 2019). As such, AI had become the center of negotiations between different societal interests.
It was in this setting that the contours of an institutionally framed European
AI agenda took shape as a distinctive cultural positioning with an emphasis on “ethical technologies” and “trustworthy AI”. It was spelled out in core documents and statements in a process that involved European member states, a European high-level expert group on AI, a multistakeholder forum called the European AI Alliance and the European Commission. EU decision-makers were recognizing that AI had become an area of strategic importance, transforming critical infrastructures in all the aforementioned sectors and was therefore also a driver of economic development. And the EU’s AI approach was on those grounds defined as a policy investment in ensuring Europe’s competitiveness on a global scale by, for example, increasing annual investments in AI development and research and establishing an agreement to join forces with national strategies on AI in member states. Thus, the AI agenda was also often described in this period as a response to a “global AI race” in the public media, debates and reports. The main focus here was the competition among regional players for global leadership on the resources for AI (e.g., data access), capital investment, AI technical innovation and practical and commercially viable research and education as well as “ethics” as a form of risk mitigation and regulation (Merz, 2019). Here, I propose that besides a race for resources, technological supremacy and risk mitigation, the explication of values-based cultural frameworks for AI played a key role.

The European Commission published its first communication on artificial intelligence in early 2018 (European Commission A, 2018), which was accompanied by a declaration of cooperation on artificial intelligence signed by 25 European member states (European Commission B, 2018) (which was later in 2018 concretized in a Coordinated plan on artificial intelligence, “made in Europe”, European Commission C, 2018). This first communication presented a general initial European approach to AI with a focus on cooperation among member states, multi-stakeholder initiatives, investment, research and technology development. Above all, AI was at this point described as part of a European economic strategy within a global competitive field. While it was not a core strategic element of this first communication on the topic, a values-based positioning was also offered: “The EU can lead the way in developing and using AI for good and for all, building on its values and its strengths.” (European Commission A, 2018) and a first step to address ethical concerns was made with the plan to draft a set of AI ethics guidelines.
Following this, a European high-level expert group on AI was established in June 2018, with 52 selected members consisting of individual experts and representatives from different stakeholder groups. Their mandate was to develop AI ethics guidelines and policy and investment recommendations for the EU. From the outset, the group’s work was framed in terms of a distinctive European framework. For example, at the group’s first meeting in Brussels in June 2018, a European Commission representative responded to a comment regarding Europe’s competitiveness: “AI cannot be imposed on us”, and it was concluded that “Europe must shape its own response to AI” [3].

Notably, the “European response” was already here defined in terms of what was presumed to be a set of shared European values. For example, at the same meeting, the chair introduced the core constituents of the group’s mandate and the European Commission’s expectations to the group as follows: “It is essential that Europe shapes AI to its own purpose and values, and creates a competitive environment for investment in AI” [4]. A decree that was later included in the discussions of the group and defined as the search for a distinctive European position in a global setting: “Discussion also centred on identifying the uniqueness of a European approach to AI, embedding European values, while at the same time identifying the need to operate successfully in a global context” [5].

The ethics guidelines published a year later in April 2019 were also outlined on the basis of “European values”. Values were introduced in this document with reference to the European Commission’s vision to, among others, ensure “an appropriate ethical and legal framework to strengthen European values” [6]. The key references here were the European legal frameworks, such as the Charter of Fundamental Rights and the General Data Protection Regulation. However, European values were also encompassed in one unifying ethics framework defined as the “human-centric approach” in which the individual human being’s interests prevail over other societal interests: “The common foundation that unites these rights can be understood as rooted in respect for human dignity—thereby reflecting what we describe as a ‘human-centric approach’ in which the human being enjoys a unique and inalienable moral status of primacy in the civil, political, economic and social fields” [7].

Yet it was the delineation of a specific type of technology design and culture of AI practitioners which in the end became the ethics guidelines’ unique
cultural positioning. Several ethics guidelines for AI had in 2019 already been published in European member states, outside Europe and by international organizations. Most notably, only a few months after the high-level expert group’s ethics guidelines were published, 42 countries adopted an Organisation for Economic Co-operation and Development (2019) recommendation that included ethical principles for trustworthy AI. Though in comparison with other more principle-based ethics guidelines, the high-level expert group’s ethics guidelines were particularly focused on the operationalization of ethics in the design of AI, that is, on framing the practice of building AI and hence providing concrete and practical guidance to AI practitioners. Europe was consequently also described in the guidelines as a potential leader in the development of “ethical technology”, with a call to create a very specific approach to the design of AI. As such, ethics and values were considered to be a property of technological design and practice, and in addition to deployers and users of AI, in the guidelines practitioners were urged to implement and apply seven ethical requirements that were supplemented with an assessment list with concrete questions to guide AI practitioners.

During the process of developing the ethics guidelines, the title of the work changed from “Trusted AI” to “Trustworthy AI” [8]. While this might be conceived of as a change primarily in semantics, the transformation in fact built on core discussions at group meetings centered on the inherent values of AI design. Accordingly, that title mirrored the conclusion of the group discussions, which was that AI technologies should not just be trusted, the EU needed to ensure that trustworthiness was built into the “technology culture” of AI innovation. As stated in the report from the first workshop of the high-level expert group, “Trusted AI is achieved not merely through regulation, but also by putting in place a human-oriented and ethical mindset by those dealing with AI, in each stage of the process” [9].

In this way, Trustworthy AI came into being as the European “third way” in AI innovation. This also meant that when working with the policy and investment recommendations that were published in June 2019, the high-level group proposed Trustworthy AI as a core European strategic area (HLEG B, 2019). Hence, the recommendations emphasized the leveraging of European “enablers” for Trustworthy AI, for example, by providing human-centric AI-based services for individuals, making use of public procurement to ensure trustworthy AI, integrating knowledge and
awareness, updating skills among policy-makers, work forces and students, developing a research university network on AI ethics and other disciplines necessary to ensure trustworthy AI across Europe, providing legal and technical support to implement trustworthy AI and mapping legal frameworks and creating new laws where the risks were considered high (e.g., when AI is used in the context of mass citizen scoring or autonomous weapons). Even recommendations were made to develop a European AI infrastructure based on personal data control and privacy (HLEG B, 2019).

Alongside the high-level expert group’s development of a set of ethics guidelines and policy and investment recommendations on AI, the way in which a European ethics and values-based approach to AI was addressed by the European Commission also transformed from a brief “concern” in a political strategy (European Commission A, 2018) into a strategic point of positioning. Nathalie Smuha, who was the coordinator of the high-level group, has described how the work of the high-level group was quickly adopted within the European Commission’s general AI strategy (Smuha, 2019). As she explains, the European Commission at that time counted around 700 active expert groups, such as the high-level expert group on AI, that were tasked with drafting opinions or reports advising the Commission on particular subjects. However, their input was not binding, and the Commission was independent in the way it took into account the groups’ advice and expertise. For example, only rarely did they become the core topic of a Commission communication [10]. Nevertheless, when the high-level expert group presented the ethics guidelines to the Commission in March 2019, an almost immediate agreement was reached to publish the last communication in the two-year period — “Building trust in human-centric AI” (European Commission, D, 2019) that stated its support for the seven key requirements of the guideline and encouraged all stakeholders to implement them when developing, deploying or using an AI system [11].

This culminated in the promise of a new president of the European Commission, Ursula von der Leyen at the end of 2019: “In my first 100 days in office, I will put forward legislation for a coordinated European approach on the human and ethical implications of Artificial Intelligence” [12].

Culture and technological change
How can we explain a forceful explication of cultural values as a strategic
interest in the face of technological change? Early in the history of the introduction of computers in society, one of the pioneers within applied computer ethics, James H. Moor, described in his famous essay “What is computer ethics?” the policy vacuums that emerge when policies clash with technological developments that force us to “discover and make explicit what our value preferences are” [13]. He predicted that a computer revolution of society would happen in two stages marked by the questions we ask. In the first “introduction stage” we will ask the functional questions — how well does this and that technology function for its purpose? In the second “permeation stage”, when institutions and activities are transformed, we will start asking questions regarding the nature and value of things [14]. The historian of technology Thomas P. Hughes similarly detailed the general developmental phases of large evolving and expanding technological systems from invention, development, innovation, transfer and growth, to competition and consolidation (Hughes, 1987, 1983). Hughes refers to “a battle of the systems” in which an old and a new system exists at the same time in a relationship of “dialectical tension” [15]. The phase of competition and consolidation is therefore also a moment of conflict and resolution not only among engineers but also in politics and law [16]. In these moments of conflict, critical problems are exposed, different interests are negotiated and finally gathered around solutions to direct the evolution of the systems. A new system, or the transformation of the old system, then evolves out of the very problems identified and solved in this phase. Unlike Moor, Hughes does not consider these moments of explication as solely induced by the transformative character of the technological systems. He considers their negotiation in complex social spaces. In fact, he holds that technologies themselves are intertwined with social, economic and cultural problems [17]. That is; in an STS perspective on technological change, such as Hughes', large technical systems are sociotechnical, meaning that they are not just material and technical but also represent complex power dynamics between multiple actors and societal interests. Therefore, they cannot be explained with a focus on technical innovation or the engineering of materials only, as they are integrally part of society at large.

As follows, to explain the socio-technical shape of the AI momentum of 2018–2019, we need to consider it as something more than just technically innovative, practically implementable and economically viable. We may
describe it as “cultural”. To do this, we need some additional perspectives. In a cultural studies perspective, culture is not just one facet but multifaceted — informally and formally created by and in interaction with people and artefacts — and the meaning of these cultural relations are in constant contestation and social negotiation. The founding Marxist scholar of the British cultural studies tradition, Raymond Williams, for example famously defined culture as “shapes”, a set of “purposes” and “meanings” that are expressed “in institutions, and in arts and learning” and in “ordinary” practice [18]. Accordingly, culture is “a whole way of life” [19]. It consists of prescribed dominant meanings and, more importantly, also the negotiations of these. The meaning of culture is in “(...) active debate and amendment under the pressures of experience, contact and discovery” [20], and as such it is simultaneously “traditional” and “creative”. Hence, there are two sides to culture: “(...) the known meanings and directions, which its members are trained to; the new observations and meanings, which are offered and tested.” [21]. In this perspective, culture is a site of power negotiation.

We may continue here and think of cultural power negotiations in the context of technological development and innovation. Here, culture, or the “cultural”, can be traced in the very design of technology. Hughes defines technological culture as a complex composite of socially embedded interests, goals and intentions [22]. Famously he held that technological systems do not become autonomous by themselves but require momentum, which depends on the interests (the culture) of the organizations and people invested in the system [23]. He mentions a few of these that were invested in the development of the modern electric power system that we might also recognize as stakeholders in the AI momentum of the 2010s: “Manufacturing corporations, public and private utilities, industrial and government research laboratories, investment and banking houses, sections of technical and scientific societies, departments in educational institutions, and regulatory bodies ...” [24]. He contends that differences in “technological styles” became particularly apparent in the twentieth century due to the increasing availability of “international pools of technology” (including, e.g., international trade, patent circulation, the migration of experts, technology transfer agreements and other forms of knowledge exchange) [25]. Accordingly, he argues that technological style is the language of culture, so to speak, or it is, as he says, an “adaption to
environment” [26]; that is to say, culture is the sum of “systemized knowledge” created in interaction with the economic and social institutions involved.

This view is characteristic of an STS perspective on the cultural components of technology development. Here, I will not get into debates regarding culture as the epistemological weight on the scale of social constructivism and relativism or technological determinism and natural realism in studies of science and technology (e.g., as represented in the debate between Callon and Latour [1992] and Collins and Yearley [1992]). That is, although I recognize that culture is a contested concept, it is more generally in STS related to the way we get to know things and the skills and resources we use to create a technology. We might say that distinct “knowledge cultures” or “technological cultures” are the foundations of a technology’s design and adaption in society. Andrew Pickering, for example, describes culture as the resources that scientists use in their work or a shared conceptual field [27].

Harry M. Collins defines cultural skills as intents and purposes and sets of rules of action for the design of a technology [28]. They are the inexplicable or “hidden” components of technology development [29]. He also argues that these implicit cultural skills of technology practitioners transform when they are made explicit and that this transformation of skills depends on changes in a “cultural ambience” that is “enmeshed in wider social and political affairs” [30].

The concept “data cultures” can here be used to illustrate the cultural variations of the different technological “styles” of the way in which data is managed and treated in technology design. These various styles could be described as the “technological cultures” of data design based on shared skills and knowledge frameworks for data technology practitioners, implicit, for example, in ideals about the big data value for technology development [31] and also explicitly described in data protection laws or ISO standards, such as the 27701 on how to create privacy information management systems (PIMS). Thus, the very practices of data scientists and designers can be said to be framed within specific cultural systems of meaning-making and accordingly the practice of developing a data system and design a cultural practice: “shaped by ideas about the cultivation and production of data that reflect epistemologies about, for example, ordering, classification, and standards” [32]. Accordingly, we may also argue that the very data design of a technology has cultural properties that
can be examined as a culturally coded system. For example, AI is not just “coded” data; it is data culture in code. As such, the AI system’s data design, or any data design, is culture in action. For example, as outlined by Collins (1987) in his description of cultural skills and AI, culture is in expert systems transformed into explicated categories, literally coded, and in advanced self-learning systems, it is even encoded within the systems when autonomous machine predictions and decisions are made; that is, the cultural classification of the world is actively coded and produced within the system.

To conclude, technology development is enmeshed in cultural spaces that can be depicted as the epicenter of interest negotiations. Notably, Hughes illustrated how each developmental phase of a technological system produces, a specific “culture of technology”, which is the sum of this complex set of interests. The technology culture is therefore, according to Hughes, also the basis of a momentum of a technological system, and, importantly, competing cultures must convert to the dominant culture of the momentum or perish (Hughes, 1987).

Making the invisible visible
Opacity was in the late 2010s often described as a core ethical challenge of the very design of AI (Burrell, 2016), on account of either intentional acts of creating obscurity with “secret algorithms” (Pasquale, 2015), unconceivable “math” (O’Neil, 2016) or permeating discursive power that concealed the interests of institutions and corporations (Zuboff, 2015). This is a core challenge that we may address here. As disparate as they may seem in their perception of the relation between culture and technological change, there is, respectively, in applied computer ethics, STS and cultural studies, a shared emphasis on the importance of making the invisible visible and explicating cultural components in order to effect change.

James H. Moor considers the “invisibility factor” [33], such as “invisible programming values” [34], a principal ethical challenge of the computer and its use per se. Collins [35] explains the move of the taken-for-granted cultural skills from inexplicable to explicable categories as a way, among others, to reduce ambiguity in knowledge and practice due to cultural and contextual distance. Hughes [36] takes a grander view when looking at the consolidation in society of larger technological systems, arguing that they do have a direction, and therefore the explication of goals is more
important for a young system than for an old one.

In cultural studies and critical data studies, the explication of cultural components is coupled with the exposure of cultural power dynamics. For example, a distinct field of feminist technoscience scholars, such as Judith Butler, Donna Harraway and Sandra Harding, have raised feminist critiques of science, technology, practices and knowledge in terms of the cultural gender power dynamics they reproduce and enforce (Åsberg and Lykke, 2010). Likewise, the data scientists and feminists Catherine D'Ignazio and Laura F. Klein (2020) describe what they refer to as “oppressive” data science cultures in their book Data feminism. These, they argue, are reflected in the goals and priorities set for the very data design of the technology. For example when minority groups are either underrepresented in data used as the basis for decisions made on social benefits, when scientific critical medical analysis only benefits one privileged group, or on the other hand when a minority group is overrepresented in data that puts them at a disadvantage in society, such as e.g., data from specific city zones used for predictive policing. In these perspectives, the various meaning making cultural practices and shared taken-for-granted cultural systems that naturalize specific situated views of the world and enforce power dynamics can only be challenged if explicated.

In other words, the cultural foundation of a technological system, what we have also referred to here as its “shape”, the “knowledge culture” behind, its “technological style”, we may also see as a prioritization of inherent values of a cultural system. In an applied ethics perspective, values are, for example, described by the philosopher of technology Philip Brey as “idealized qualities or conditions in the world that people find good” [37]. These are ideals that we can work towards realizing in the design of a computer technology. Thus, technologies can have a specific cultural shape that consists of the implicit systems of organized knowledge, practices and meanings that go into their design. Values are not just personal ideals or transcendentally “true” or “good”; they are culturally situated and constantly engage with shared cultural purposes and common meanings by enforcement and/or negotiation (Williams, 1993). This is also true for our ethical thinking about digital technologies where culture, as for example Charles Ess has illustrated in his analyses of ethics, culture and technologies, plays an essential role. Accordingly, in Western societies an
ethical emphasis has been placed on “the individual as the primary agent of ethical reflection and action, especially as reinforced by Western notions of individual rights” [38]. As such, culturally situated ethical thinking also has an interest in the power dynamics of society regarding who or what ethics is for.

In this line of argument, a first step to guide the development of trustworthy AI would be to make the cultural foundation (the data cultures) visible. Essentially we need to consider this explication of cultural components as an ethical and moral choice. As the information studies scholars Geoffrey C. Bowker and Susan Leigh Star [39] state in their work on classifications and standards in the development of information infrastructures, “Each standard and each category valorizes some point of view and silences another. This is not inherently a bad thing — indeed it is inescapable. But it is an ethical choice, and as such it is dangerous — not bad, but dangerous.”

Data interest analysis: The European agenda’s cultural shape of AI
I have so far examined how a European AI Agenda evolved over a two-year period into a distinctive European cultural positioning with an emphasis on “ethical technologies” and “trustworthy AI” [40]. First and foremost, I examined this as an interest in shaping a technological AI momentum. In this last part of the article, I move on to an investigation of four cultural components of this cultural interest in the data of AI as it was explicated in an institutionally framed process between 2018–2019.

The four cultural components of the European data interest in AI
As illustrated in the two-year period, a negotiation of a shared cultural framework for the development and adoption of AI took place and was above all broadly defined in terms of European values and ethics. Importantly, this also included a conceptualization of a European technology culture. I propose here that the European AI agenda sought to explicate this in four cultural components: 1. the cultural context, 2. the cultural foundation, 3. the technological data culture and 4. the cultural data space.

1. The cultural context: Defining the technological momentum
As we have learned, a technological system does not evolve autonomously, it is directed within a momentum that arises from the interests invested in
the system (Hughes, 1983). The culture of a larger technological system is internal to the system in the sense that it represents the sum of the focused interests and forces at play in the momentum of this particular system. But culture is also a force external to the very system, a “cultural ambience” that is entangled in general social and political affairs [41]. Transformations in the resources, skills and knowledge that drive the development and adoption of a technological system can therefore also be influenced by changes in this “cultural ambience”.

Although AI systems had already been adopted and integrated primarily in some parts of the private sector in the late 2010s, their general adoption in European society, including in the public sector was a recent development, and in policymaking, the forceful focus on AI was new. Therefore, we may consider AI in terms of what Hughes (1983) refers to as a “young system” in society in which the explication of goals is particularly prevalent. Along these lines, the European AI agenda may equally be considered a cultural interest in shaping the technological momentum of AI systems and directing their evolution in society in Europe and globally. The high-level group’s policy and investment recommendations (HLEG B, 2019) published one year into the period in which the European AI agenda unfolded describe the different societal phases of digitalization where AI forms a “third wave” characterized by its adoption in European society: “Europe is entering the third wave of digitalization, but the adoption of AI technologies is still in its infancy. The first wave involved primarily connection and networking technology adoption, while the second wave was driven by the age of big data. The third wave is characterized by the adoption of AI which, on average, could boost growth in European economic activity by close to 20 percent by 2030. In turn, this will create a foundation for a higher quality of life, new employment opportunities, better services, as well as new and more sustainable business models and opportunities” [42].

The two-year period was characterized by a sense of urgency to gain force within a global AI momentum, and in particular the stakeholders that make a momentum were therefore a central topic of the negotiation and debate. This included, for example, a focus on AI practitioners, entrepreneurs, data analysts, educators, the work force, policy-makers and citizens in general. Not only were the stakeholder interests of the members of the high-level expert group a continuous topic of contestation in public debate, but
generally a broad range of societal stakeholders were either sought out to participate in, for example, the AI alliance multistakeholder online platform created as part of the strategy and the public consultations of the high-level expert group reports, or they were addressed in the content of the reports and in presentations at various public events.

The depiction of an AI momentum was prevalent at the first public event that the high-level expert group was invited to attend [43]. Launched with a press release emphasizing the role of AI in “boosting European competitiveness” [44], the event started off with a speech by the then Commissioner for European Commission for Digital Economy and Society Maryia Gabriel who outlined the strategic goals of the European Commission with a clear message. European stakeholders could indeed shape the direction of AI: “We all have an important role to play in defining a shared European vision for Artificial Intelligence. Yes, ladies and gentlemen, digitalization is everywhere, data is everywhere. This is just the beginning of a new technological revolution.” [45]

Notably, like many others in this period, Gabriel in her speech also equated digitalization with data and at the same time described data as the foundation for AI evolution. In fact, the first European Commission Communication on AI had recognized data as a key factor for the development of AI in Europe with a reference to the creation of “data rich environments” as “AI needs vast amounts of data to be developed” (European Commission A, 2018). Thus, the main driver for AI was held to be data. As described in the high-level expert group’s policy and investment recommendations, this “third wave” of technological development was in fact “driven by the age of big data”. The EU was here consequently also described as “a pivotal player in the data economy” [46] as data “is an indispensable raw material for developing AI” [47]. Therefore, data was also held to be core to what the stakeholder interests of the AI momentum were invested in: “Ensuring that individuals and societies, industry, the public sector as well as research and academia in Europe can benefit from this strategic resource is critical, as the overwhelming majority of recent advances in AI stem from deep learning on big data” [48].

2. The cultural foundation: The values and ethics framework

Culture is a shared conceptual framework for meaning production that
consists of what we know and what we are trained in. A conceptual values-based framework for personal data is, in the European context for example, formalized in legal frameworks, such as the General Data Protection Regulation and the Charter of Fundamental Rights. But culture also consists of the new meanings that are offered and contested meanings (Williams, 1993); that is to say, culture is also a cultural negotiation in which different cultures clash and conflicts of interests may emerge.

With the rise of data intensive technologies, such as AI, not only the law but also the meaning of a traditional European approach to handling personal data was challenged, and a process of cultural meaning negotiation was therefore initiated. This we may refer to as “data ethics spaces of negotiation” (Hasselbalch, 2019) that exposed the cultural contexts that were shaping the ethical thinking of this period and ultimately sought to resolve conflicts between value systems in conflict.

As described, the European AI agenda explicated a general human-centric approach that stressed that human interest prevailed over other interests as well as a particular approach to data governance that emphasized the empowerment of individuals in the handling of their personal data. For example, the high-level expert group’s ethics guidelines outlined a clear framework for the management of data with one of the seven requirements, “privacy and data governance”, specifically addressing the human-centric values embedded in the data design of an AI technology. In this context, the concept of human agency stood out as the individual’s knowledge and the information provided for the individual to make decisions and challenge automatic systems [49].

The human-centric approach came to represent the overarching framework of the European AI agenda for resolving different interests and values embedded in AI innovation. Conflicts existed between data protection/privacy, ethics and data-driven innovation, machine automation and the human work force, the interests of the individual and society/public institutions as well as scientific and governmental interests. That is most important as stated in the policy and investment recommendations: “AI is not an end in itself, but a means to enhance human well-being and flourishing” [50]. As a result, human-centric practical solutions for resolving such conflicts were suggested: ethical technology as a competitive advantage (resolving conflicts between ethics and data-driven innovation), humans-in-the-loop AI solutions for the work
place and upscaling the AI skills of the work force (resolving conflicts between automation and the replacement of workers), data design as an enabler of human well-being and protection such as developing mechanisms for the protection of personal data and individuals to control and be empowered by their data (the interests of the individual and society) and generally focusing on the use of non-personal data in business to business (B2B) AI solutions rather than the personal data of B2C solutions (resolving conflicts between risks of using personal data and the data intensity of AI technology development).

3. The technological data culture: Skills, knowledge, style and resources

Technological development is not neutral. Engineers and designers develop technologies within shared knowledge cultures that form the foundation for their work. These foundational cultural frameworks can be described as “technology cultures” — shared fields of resources, implicit and/or explicit skills, experiences, methods and even tools they use when they build technologies and that therefore also contribute to the shaping of technological development.

As described previously, during the process of developing the European AI agenda, the explication of a European “technological culture” for the development of AI became an essential focal point. In this respect, the skills, the education, the methods and practices needed in the developmental phase of what was referred to as “ethical technology” was core to discussions concerning economic investments, awareness raising and policies. In fact, as previously illustrated, a European ethical design culture grew into being as the European position of the global AI momentum.

In 2019 at the first AI Alliance assembly in Brussels, Commissioner Maryia Gabriel talked about getting the “policy right”, which meant adopting and developing AI with “a decisive, yes, but” as she said. This “but” was a reference to the European risk mitigation of the ethical challenges of AI [51]. The first, she mentioned, was global competition (e.g., that the EU was several billion euros behind in terms of investments in AI), the third was “ethical and legal concerns”. In between was the second challenge, which according to Gabriel was the social impact of AI. Her suggestion was to invest in education and training, digital education plans and developing digital skills in Europe.
The European strategic investment in a particular “technology culture” of AI was also an essential focus of the high-level expert group’s policy and investment recommendations. It first and foremost came to equate a shared foundational AI knowledge culture. Europe needed to “foster understanding” and “creativity” [52] and generally “empower humans by increasing knowledge and awareness of AI” [53]. In this way, an entire section of the recommendations focused on “generating appropriate skills and education for AI.” This was not just limited to technical skills, but also “socio-cultural skills” [54]. In general, there was a key focus on the development of new skills or the updating of skills of not just engineers but also policy-makers and the general work force. This was also extended with a call to develop basic education on AI and literacy in higher and lower education and an “AI competence framework for individuals” [55]. In many instances, the term “data literacy” was here used interchangeably with the concept “digital literacy”. Markedly, the public sector was described as playing a fundamental role in the development of a Trustworthy AI “technology culture”, e.g., by fostering “responsible innovation” through public procurement.

One thing was the assumption that a particularly European “technology culture” of AI was needed for Europe to succeed in global competition. But how was this “technology culture” then explicated? Here, the assessment list of the ethics guidelines was particularly interesting as it explicated in detail concrete questions to guide the design, management and development of AI within each of the seven requirements for trustworthy AI. We saw here explained a “data culture” for AI, most explicitly in section 3 on “Privacy and data governance”. The point of departure was privacy and data protection, and thereafter we moved on to ensuring the quality and integrity of data and procedures for managing access to data.

4. The cultural data space: The infrastructure

According to Hughes (1983), technological style differs from region to region and nation to nation. He equates culture with geographically and jurisdictionally delineated spaces. But we may also add to this depiction the technological evolution of space that has challenged this very correlation between culture, geography and jurisdiction. As a consequence, culture is no longer just the asset of a nation, rooted in geography and national law, but is increasingly extended into virtual communities with “cultures” or
“subcultures” delineated by symbolic borders of cultural values and ideas. At the beginning of the twenty-first century, “data cultures” had been created on the basis of an interjurisdictional digital flow of data. As such, the very “architecture” of a global data infrastructure had emerged as an interjurisdictional space challenging first and foremost European data protection/privacy values and legal frameworks. For example, looking at case law of the European Court of Human Rights (ECHR), it very early started considering the level of uncertainty that the challenges of technological progress posed to the ECHR’s territorial definition of jurisdiction in cases concerning the right to privacy [56].

In the 2010s, AI was developed primarily on the basis of an interjurisdictional and territorial global big data infrastructure. However, the revelations of embedded data asymmetries in the form of surveillance scandals, fake news and voter manipulation had provoked a European concern with foreign “data cultures” and their “data architectures”. The European AI agenda proposed an alternative European data-sharing infrastructure for AI based on a foundational values-based approach to data but which was also confined within the European jurisdiction and geographical space. In the policy and investment recommendations published, the high-level group described data infrastructures as the “basic building blocks of a society supported by AI technologies”. Data infrastructures were described as the foundation of a European AI critical public infrastructure, and therefore should be treated as such: “Consider European data-sharing infrastructures as public utility infrastructures.” Thus, the development of this European space should also be invested with a specific set of values and designed “with due consideration for privacy, inclusion and accessibility, by design” [57].

It was particularly in this description of a European AI data infrastructure and architecture that the cultural interest in data stood out. Thus, the values-based approach was also conceived of as a cultural effort to transfer European values into technological development, positioned against a “non-European” threat perceived to be pervasively embedded in technological infrastructures: “Digital dependency on non-European providers and the lack of a well-performing cloud infrastructure respecting European norms and values may bear risks regarding macroeconomic, economic and security policy considerations, putting datasets and IP at risk, stifling innovation and commercial development of hardware and
Conclusion
As wild and unruly as it may seem, construed in a hodgepodge of complex relations, interests, symbolic meaning-making, people and artefacts, a technological momentum also has a shape — a shape that guides its direction, values, knowledge, resources and skills that form its technological architecture and its governance, adoption and reception in society. At times, this shape is more explicitly “cultural” and values-oriented than others when it is large and socially and culturally transformative for example or when it spreads on a global scale.
The global AI momentum of the 2010s was a moment like that. Big data systems empowered by AI technologies were transforming European societies, challenging what was held to be European fundamental values, and driving out an explication of what it means to do AI in the “European way”. With which values should AI be designed? Which interests should drive the development? What skills and education? What role should technology and science play in society? And could Europe even compete on those grounds? Information policy approaches were transforming from having narrow functional focuses on the digitalization of “everything” to more complex and multifaceted values-based emphases on the ethical and social implications of data technologies, including everything from legislative measures in competition, data protection, criminal and consumer protection law to research and innovation investments in “ethical technology” development and a European datasharing infrastructure.
In 2018, Europe had been going through a period of self-exploration regarding the role of big data and emerging technologies in European societies in general. Following an all-encompassing digitalization wave, the social and ethical implications were materializing with big data scandals and revelations. A recent reform of the European data protection legal framework was presented as Europe’s powerful global response to these challenges. However, a law did not seem to be a sufficient governance response by itself, and therefore a process was initiated to develop what was referred to as a European approach to what was perceived as a general AI evolution of the age of big data.
In this article, I have investigated a cultural interest in the global AI
momentum—the cultural shape it took with an emphasis on “ethical technology” and “trustworthy AI” in response to global AI innovation, how it evolved in a process of public events and a high-level expert group on AI established by the European Commission and how this cultural interest took form as a data interest that was sought to be explicated in policy and investment recommendations as well as in a set of ethics guidelines. I relied on the thesis that technological development is not neutral. This also means that the culture of a technological design is not a randomly adapted technological style. It is one sum of interests, value frameworks and the negotiations of these, and if these are made visible, we can argue that the technological development of society can be shaped and chosen. The choice to do AI ethically and responsibly is not a simple one; in fact, it is as complex as the culture we are trying to shape with it.

I here want to suggest that to direct the AI momentum of the age of big data, we need an ethics concerned with the embedded data interests and powers, what I have also referred to as a “data ethics of power” (Hasselbalch, 2019). I have previously (Hasselbalch, 2019) described how European policy and decision-makers in the late 2010s were positioning themselves against a threat to European values and ethics perceived to be embedded in the big data socio-technical systems of what was named “GAFA” (acronym for the four big U.S. tech companies Google, Apple, Facebook, Amazon). Considering this “cultural ambience” (Collins, 1987), I propose that the cultural positioning of the European AI agenda may also be viewed as a data ethical choice formulated in direct response to the technological data cultures of the dominant AI technologies at that time.

In the article I focused primarily on the institutional explication of European cultural values as an interest in a technological momentum. I did not seek to predict the actual adoption and implementation of AI in Europe. Nevertheless, a few considerations and recommendations can be made concerning the implementation of a European “third way” on the global arena with an emphasis on the development of trustworthy AI and the human-centric approach. Following the two-year period that I examined in this article, the European Commission in 2020 published a comprehensive strategy on the digital, AI and data future of Europe (E, F, G). While this strategy was ambitious with respect to furthering Europe’s position in the “global AI race” by advancing a general AI uptake and taking back control of a European data resource space, the values-based European
“third way” was mostly addressed in a legal compliance and requirements framework (Hasselbalch, 2020). Based on this article’s delineation of the complex factors that constitute an ethical “data culture”, we may argue that this is not enough and that an additional set of combined governance tools is needed to shape the technological AI momentum’s data cultures as “trustworthy”. For example, we need investment, innovation, research and education in “ethical technology” components and processes in specific (from human-in-the-loop features, state of the art anonymization techniques to ethical impact assessments). The ethical component of AI implementation in Europe cannot just be ancillary to the European AI uptake; an ethical data culture for Europe needs a dedicated economic, social and political investment.

Notes
1. The definition of artificial intelligence has changed throughout history since the 1950s with the development of different scientific and social paradigms. As such, in the 2010s, the term AI still did not have one shared signification. In this article, I do not consider the “intelligence” of AI (technologically or philosophically), but use the term AI generically to address public discourse on the topic. My emphasis is on AI as automated decision-making data intensive systems that are designed to perceive their environment through acquiring data and interpreting the data to decide action to achieve a goal (HLEG A, 2019, p. 36).
3. HLEG C, p. 4.
4. HLEG C, p. 2.
5. HLEG C, p. 5.
6. HLEG A, p. 4.
7. HLEG A, p. 9.
8. Upon suggestion from the author of this paper and based on a conversation with Sille Obelitz Søe. Argument for this revision can be found here: https://dataethics.eu/why-trust-in-ai-is-not-enough/.
9. HLEG D, p. 2.
40. I examine the European AI strategy (described in the two communications “Artificial intelligence for Europe” and “Building trust in human-centric artificial intelligence” (April 2019), in the “Declaration of cooperation on artificial intelligence” (April 2018) and the “Coordinated plan on artificial intelligence ‘made in Europe’” (December 2018)) with a core focus on the work of the European high-level group on AI and the two core deliverables of this group: the “Ethics guidelines for trustworthy
AI” (April 2019) and the “Policy and investment recommendations for trustworthy AI” (June 2019). The investigation is based on a qualitative reading of these documents and includes perspectives from the process of the very development of these two documents (with reference to public minutes and records from meetings) as well as from concurrent European policy responses.


42. HLEG B, 2019, pp. 6–7.

43. The AI forum held in Helsinki in October 2018, co-hosted by the Ministry of Economic Affairs and Employment of Finland and the European Commission.


46. HLEG B, 2019, p. 16.

47. HLEG B, 2019, p. 28.

48. Ibid.

49. HLEG B, 2019, p. 16.

50. HLEG B, 2019, p. 9.


52. HLEG B, 2019, p. 9.

53. HLEG B, 2019, p. 10.

54. HLEG B, 2019, p. 32.

55. HLEG B, 2019, p. 10.


57. HLEG B, 2019, p. 28.

58. HLEG B, 2019, p. 3.

References


292


European AI agenda document


Abstract
This article makes a case for a data interest analysis of artificial intelligence (AI) that explores how different interests in data are empowered or disempowered by design. The article uses the EU High-Level Expert Group on AI’s Ethics Guidelines for Trustworthy AI as an applied ethics approach to data interests with a human-centric ethical governance framework and accordingly suggests questions that will help rebalance conflicts between the data interest of the human being and other interests in AI design.

Introduction
Data flows. Data transforms. Data changes hands, bodies and containers. The next frontier in the age of big data flows is artificial intelligence (AI) systems that are developed to contain and make sense of large amounts of data and to act on that knowledge. As AI data systems become integrated into society, they also turn into centers of negotiations between different interests in data that join, appear and disperse in the data design of AI as a symptom of and condition for the distribution of agency and powers in society. This article argues for an exploration of data interests in AI to help make ethical choices in the development of AI.

In the late 2010s, the term “artificial intelligence” gained traction in public discourse. Business and technology companies started rebranding their big data efforts as “AI” (Elish & Boyd, 2018), and in policymaking AI became an item of strategic importance worldwide. In public and private sectors, decision-making processes were progressively informed by and even replaced by big data AI systems. Recommendation and personalization systems were profiling and analyzing our personal data and deciding for people what they see and read and with whom they engage online. Judicial risk assessment systems were looking for patterns
in backgrounds of defendants to inform judges about who would be most likely to commit a crime in the future. Triage systems processed the medical and demographic history of patients to decide who would get a kidney. Moral decisions and choices were increasingly intertwined with AI systems’ complex data processing, and accordingly interests in the data of AI as a resource to protect, share, acquire and to be empowered or disempowered by came together in efforts to direct the development of AI.

Against this background, an institutionally framed European AI agenda took shape with an emphasis on “ethical technologies” and “Trustworthy AI”. This applied AI ethics approach was described in core documents and statements in a process that involved, among others, an independent multi-stakeholder expert group on AI (“The EU High-Level Expert Group on AI”, AI HLEG) that in 2019 published the “Ethics Guidelines for Trustworthy AI”.

Several other ethics guidelines and sets of principles for AI, such as these, were at the time created by various stakeholder interest groups. Most of them have been shown to share common themes (Fjeld et al., 2019; Floridi et al., 2018; Jobin et al., 2019; Winfield & Jirotka, 2018). Yet, while thematic convergence is indeed relevant when seeking out general global consensus on applied ethics approaches to AI, I want to argue here that each guideline also has a unique point of reference and, importantly, is positioned within distinct cultural, social and legal frameworks.

The AI HLEG guidelines were framed with a human-centric approach to the development of AI: “The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored…” (p. 37). In the late 2010s this “human-centric” approach was a term and a theme that had emerged in policy discourse on AI with no common conceptualization other than an emphasis on the special role and status of humans (e.g. OECD, 2019, European Parliament, 2019). However, the AI HLEG’s work with the ethics guidelines was from the outset framed in terms of a distinctive European agenda created to ensure the development of a European AI eco system (Hasselbalch, forthcoming). The guidelines’ “human-centric” approach was therefore also grounded in a European fundamental rights legal framework and the European data protection legal reform with an emphasis on the autonomy and dignity of the individual human being.
The idea that the design of a socio-technical system may be informed by overarching ethical reflection, such as the human-centric one, calls attention to the moral qualities of technologies, that is, respectively their embedded “values” or their “politics”. In the following, I describe AI systems as agents to which humans delegate the enforcement of the agency of different interests in society. Recognizing this type of delegated agency of interests allows us to develop and design with a human-centric ethical reflection and to make choices accordingly. I begin with an examination of the concept of “the human interest” and “data interests”. The first premise for a data interest analysis is that the very design of a data system represents power dynamics that a technology may create by design “wonderful breakthroughs by some social interests and crushing setbacks by others” (Winner, 1980, p. 125). While an applied ethics approach, such as Value Sensitive Design (VSD), allows us to isolate the very design phase of a data system to analyze the way in which stakeholder values are negotiated by design, a Science of Technology (STS) framework treats the power (moral or political qualities) of technologies as dynamic concepts in constant negotiation with societal factors. This is the second premise of the article that considers the development of AI data systems as a component of society at large, evolving laws and policies, economy and culture. Here, I move on to the depiction of data interests in the AI HLEG’s ethics guidelines. As the AI HLEG does not explicitly address data interests as such, I use a metaphorical analysis as a tool to illustrate the guidelines’ position on how to resolve conflicts between different data interests. I suggest five themes in regard to data interests (“data as resource”, “data as power”, “data as regulator”, “data as vision” and “data as risk”) that should be considered in the ethical governance of AI.

This article is not concerned with the philosophical concept of AI, nor with the scientific aspiration to create human-level machine intelligence. It addresses the practical adoption of AI in society as data intensive systems and uses a specific applied ethics approach to the development of these systems as suggested by the AI HLEG. Therefore, the focal point is, above all, on machine learning, which is essentially a data processing system with different degrees of autonomy. By the late 2010s, amplified computer power and the vast amount of data generated in society had empowered machine-learning technologies to evolve and learn to recognize faces in pictures, to recognize speech from audio, to drive a
car autonomously and to understand individuals when, for example, microtargeting services and information, etc. These were all practically applied, more or less autonomous systems of data processing adopted by companies and states to not only solve simple problems, analyze and streamline disparate data sets but to act in real time, sensing an immediate environment and to support critical human decision-making processes. As such, their ethical implications in regard to human agency and involvement arose from systems of distributed moral decision-making between humans and “non-human” agents.

The Human Interest in AI development

The human-centric approach to AI emerged as a shared emphasis in the policy debate on AI of the late 2010s and is generally associated with the well-being of the individual human being and the societies and environments of humans in the design, development and implementation of AI of (e.g. in AI HLEG B, 2019, p.9). As follows, a key emphasis is placed on AI design that prioritizes human involvement, human agency and human decision-making. In the European policy context, this is first and foremost associated with the autonomy, dignity and fundamental rights of the individual human being. As core concepts of the European Fundamental Rights framework that form part of the core principles of European law, the human-centric approach therefore here also goes beyond a concern with the design of AI systems only, compelling a socio-technical environment in which fundamental rights are implementable.

AI systems have been deployed in different societal sectors since the 1980s. They evolved from rule-based expert systems encoded with the knowledge of human experts that were applied in primarily human and physical environments into machine-learning systems, evolving and learning from big data in digital environments with increasingly autonomous decision-making agency and capabilities (Alpaydin, 2016). Despite the fact that machine learning to a certain extend cut out the human expert, it did not entirely exclude human involvement in the very data design of the system. On the contrary, the degree of an AI system’s autonomy is, for example, prescribed by the human involvement in data processing from the definition of the problem, the collection of data and data cleaning to the training of the machine-learning algorithm (Lehr & Ohm, 2017). Lehr and Ohm refer to this human involvement in machine-
learning processes as “playing with data”. Machine-learning algorithms, as they state, are not “magical” black boxes with mysterious inner workings. In fact, they are the “complicated outputs of intense human labor—labor from data, scientists, statisticians, analysts, and computer programmers” (Lehr & Ohm, 2017, p. 717).

In this way we may associate the human interest in the data of AI in very practical terms as an involvement of human actors in the very data design, use and implementation of AI. In the AI HLEGs ethics guidelines the human-centric approach is for example spelled out in a particular attention to the interests of the individual human being and “human-in-command” and “human agency and oversight” components in the design and conditions for the development of AI. Though the preoccupation with the human interest in AI data design may also be considered more generally in terms of a critical reflection on the very status of human agency and AI agency. Responding to a conflict in the original aspirations of AI research to create respectively machines that thinks and understands by themselves or machines that “just” process information and solve problems for humans, John R. Searle famously argued in 1980 that strong AI has “little to tell us about thinking, since it is not about machines but about programs, and no program by itself is sufficient for thinking.” (Searle, 1980, p.417). Today public discourse on AI carries traces of the first aspiration to build artificially intelligent agents comparable to the human agent and accordingly conceptions of the imminent potentials or threats of AI. For example, concerns regarding AI agents that replace the human labour force or the artistry and creativity of new AI systems represent an imagining of an autonomous out of human control artificial agent. As such, it may be argued as, Ellish & Boyd does, that this type of “magic” surrounding Artificial Intelligence also disempowers us in what we think we can do with AI (Ellish & Boyd, 2017).

Along these lines, another “human-centric” call for action has been made by Spiekerman et al. (2017) in their “Anti-Transhumanist Manifesto” that directly opposes a vision of the human as merely information objects no different than other information objects (non-human agents) which they among others describe as “an expression of the desire to control through calculation” (p. 2). In this way, the human-centric approach can therefore also be depicted as a more general concern with the agency of humans in the control structures of their technical environments (Deleuze, 1992)
Building on this conception of a human-centric approach, I in the following purposefully make a distinction between the “human” and “non-human agent”. In other words, AI can be defined as complex socio-technical data systems invested with interests and providing agency to different stakeholder group interests in society, but at the same time the very conception of the autonomous agency of AI can be considered an interest to be harnessed. As such, a human-centric distribution of power may be operationalized with design components of AI that enable the agency and critical reflection of human beings in semi-autonomous systems, and we may therefore also argue that it is this very human component of the AI system’s design and use that should be the focus of a human-centric ethical governance approach to AI.

**Data Interests by Design**

Technological design is a dynamic process that is simultaneously shaped by society and is society-shaping (Hughes, 1987, p. 51). It embodies a range of diverse factors technological, social, political, economic, as well as the individual professional skills and prejudices of engineers that are all “thrown into the melting pot whenever an artifact is designed or built” (Bijker & Law, 1992, p. 3). Nevertheless, as Bijker and Law (1992) put it, technologies do not represent their own inner logic; they are shaped, sometimes even “pressed”, into a certain form that “might have been otherwise” (p. 3). From this perspective, we may consider the embedded data interests in the very technology design as a component in a network of factors that to a certain extend can be “ethically” guided. This very focus on design as one component of ethical governance may also be referred to as an “ethics by design” approach that aims to develop methods and tools for designing ethical behavior in autonomous agents to warrant that they behave within “given moral bounds” (Dignum et al., 2018, p. 61).

The idea that human values intentionally can be designed into a computer technology was originally formulated by Batya Friedmann and partners in the 1990s and has since then been further explored in the Value Sensitive Design (VSD) applied ethics framework (Friedman, 1996; Friedman et al., 2006; Friedman & Nissenbaum, 1995, 1997). In VSD, the embedded values of a technology are addressed as ethical dilemmas or moral problems to solve in the very design of computer technologies. A data interest may, in this context, be described as an intention or a motive
that is transformed into specific properties of a data technology that arranges data in ways that support the agency of certain interests in the data that is stored, processed and analyzed by the AI technology. An applied ethics VSD approach would here aim to solve conflicts of data interests in the very design of a technology. For example, one could consider the individual’s interest in “privacy” as a value that is adversely affected by a specific design of a data-intensive technology, and accordingly suggest an alternative design in which privacy is deliberately designed into that technology (for example with a “privacy-by-design” [Cavoukian, 2009] approach). In VSD, interests are correlated with values that are held by stakeholders and that can be affected adversely or supported by a technology’s design. Thus, the aim is to develop analytical frameworks and methodologies to rebalance the distribution of interests and attempt at resolving conflicts between the needs and values held by these stakeholders in the design of a computer technology (Umbrello, 2019; Umbrello & De Bellis, 2018). These stakeholder analyses may also be extended to policy contexts in which technology design is negotiated. Steven Umbrello (2019, p. 7) for example identified specific values (data privacy, accessibility, responsibility, accountability, transparency, explainability, efficiency, consent, inclusivity, diversity, security, control) in the committee evidence reports of the UKs Select Committee on Artificial Intelligence tracing them directly to the different stakeholder groups involved in the committee (Academics, non-profits, governmental bodies and Industry/for profits) ranking their order of distribution).

Data Interests in Society

Data interests can be explicitly examined during the different design phases and the deployment of AI. Every day negotiations are taking place among different interests in society in very tangible digital data resources. These data interest negotiations we see represented in concrete deployments of AI technologies. They represent micro individual stakeholder objectives, values and needs, from developers to the users, institutional or business interests in data or they may even represent macro cultural sentiments or social structural requirements. Data interests may be aligned or compete, some are explicitly considered, but many are not considered at all in the developmental process or deployment of AI. An example is the “babysitter app” Predictim that uses language-processing
algorithms and image-recognition AI software to analyze babysitter job seekers’ social media posts to produce a personality report with a risk score (Harwell, 2018). At first glance, the app developers have an interest in bettering the AI functionality by enriching it with data from social media. From the users of the app’s point of view, they have (as employers) an interest in the insights provided by the data analysis of the app. The job seekers on the other hand might have an interest in keeping their social media data private or in just reviewing the report based on their data, which is only accessible to employers. Facebook and Twitter also had an interest in the social media data processed by the app and therefore banned the app from their portals. Predictim’s CEO said the company was not doing anything wrong: “Everyone looks people up on social media, they look people up on Google” (Lee, 2018). What he did not consider was a shift in the general societal interest in the data of AI; that is, an increasing concern with the social implications of big data technologies.

As a classic sociological concept, interests are considered major determinants of social action (Spillman & Strand, 2013, p. 86). Interests are sustained and amplified in social processes that determine the power dynamics in society. Recent studies from different scholarly fields have illustrated concrete examples of the social implications of AI, highlighting the power relations and interests at play in the development and societal adoption of AI. Data systems and mathematically designed algorithms are not impartial or objective representations of the world but are invested with the interests of the powerful—governments, public institutions and big data industries. They therefore trigger actions with ethical and social implications.

Examples of the social and ethical implications of AI and data intensive systems in general are manyfold. The mathematician Cathy O’Neil (2016) describes what she refers to as “weapons of math destruction” (WMDs) that are deployed by private and state actors without question as neutral and objective systems replacing human decision-making and assessment with devastating consequences for citizens. A teacher loses her job due to a rigid machine-based performance assessment that does not take into account social contexts and human factors; a young person from a rough neighbourhood gets a visit from the police based on their use of a predictive crime tool. The legal scholar Frank Pasquale (2015) equally worries about the increasing use of automated processes to assess risks and
allocate opportunities. These, he argues, are controlled by private companies that are also the most profitable and essential parts of the information economy. Another famous case study of the power relations at play in the adoption of AI systems is the “machine bias” study published by the news site ProPublica. Here, the investigative journalist Julia Angwin (2016) together with a team of journalists and data scientists examined the private company Northpoint’s COMPAS algorithm, which is used to perform risk assessments of defendants in the U.S. judicial system and assess the likelihood of recidivism after release. They found a bias against black defendants in the algorithm that had the tendency to designate black defendants as possible reoffenders twice as much as it did with white defendants. At the same time, it more often grouped white, rather than black, defendants as a low risk.

The Delegated Agency of AI

Langdon Winner (1980) describes technologies as “ways of building orders in the world”, as active “structuring activities” by which “different people are differently situated and possess unequal degrees of power as well as unequal levels of awareness” (p. 127). Technologies are not neutral tools but have embedded politics, he argues. They are the locus of the distribution of societal powers shaped by human motives, and therefore they also embody “power and authority” (Winner, 1980, p. 131).

In a big data socio-technical environment, we may also see ourselves as locked in specific positions prescribed by the “politics” of different combinations of data sets that either grant us access or restrict access to insights and connections that can be transformed into lost or found opportunities. Bolukbasi et al. (2016), for example, examined how gender bias in data from news articles is reproduced and amplified in some of the most popular machine-learning methods for language processing that are used in online search engines. They found that due to existing bias in the training data of the model, words are organized in clusters of words such as architect, philosopher, financier and similar titles grouped together semantically as “extreme he” words, whilst words such as receptionist, housekeeper and nanny are grouped together as “extreme she” words (Bolukbasi et al., 2016, p. 2). An employer’s online search for a candidate for a job position, might therefore very well present to her a list of information embedded with this type of pre-existing societal bias. This list
represents the delegated moral agency of humans and society, which prescribes a very specific prioritization of the information she should look into. Whether she acts on the information she accesses based on this list or not, the design of it can still be argued to have embedded “politics” and “values” that are inscribed in the algorithm of the search engine. Thus, we may argue that choices in regard to the very data design of an AI system have potential implications for the agency of different data interests.

Deploying a perspective from economic theories on agency, we may here consider the informational relations between “agents” that manage the (social and/or economic) interests of “principals”. To provide an example, an agent could be a real estate agent, and the principal could be the buyer and/or seller of a property whose primarily economic interests are negotiated in the information contained and shared in contracts and relationships. Applying this perspective in social theory, the fundamental idea is that we very often in society do not negotiate our own interests. We have allocated agents that negotiate on our behalf (Shapiro, 2005). That is, we delegate decision-making authority to agents that are to represent our interests. Thus, an agent could for example also be a nation state, and the principals could be the citizens whose agency is affected by the informational relationships they have with their government.

The concept of “informational asymmetries” between principals and agents and consequently trade-offs between interests is of core importance to a critical analysis of agency (Shapiro, 2005). A state’s unaccounted-for mass surveillance of citizens constitutes an informational asymmetry; so too does a social media company’s opaque harvesting and processing of the personal data of its users. Both forms of asymmetry have a direct effect on citizens’ and individuals’ agency.

The idea that our agency is defined by the informational relationships we have with different types of economic, social or political representatives that negotiate interests on our behalf is relevant to a data interest analysis of AI. That is, increasingly we delegate decision-making to AI technologies with built in trade-offs between different interests that might be correlated with or in conflict with our own interests. Hence, we may consider AI technologies as “moral agents”, not with moral agency as such (Adam, 2008) and capable of their own of making moral decisions but as agents to which humans delegate the enforcement of the agency of different interests (Latour, 1992). AI technologies can in this way be
described as agents representing our interests and acting on our behalf. They are nodes that manage informational relationships between different societal actors, distribute informational resources and allocate agency of interests.

To this end, the very design of the informational relationship (the data design) we have with our AI agents, the insight and access to data, constitutes the balancing and trade-offs between interests. Data design might embody “informational asymmetries”, the implicit other of what in economic agency and game theory is referred to and aspired to as a desired state of “perfect information” in which all interests are served by having an equal amount of information to make rational decisions (Von Neumann & Morgenstern, 1953). Of course, technological development is always embedded with societal trade-offs and consequently ethical choice. Information is never perfect. It has a dimension of moral choice that entails trade-offs between interests, and it is this choice that requires an ethical governance framework.

**Ethical Governance**

If we consider AI data design as a type of agent with a particular organization of data interests that supports or represses their respective agencies, a core question to pose to the development of AI is if an alternative data design would balance out conflicts of interests in a human-centric framework to the benefit of the individual human being. The idea that we can design and govern technologies with an aim as such in mind, or an “ethics by design” approach, calls for an emphasis on the design of the moral qualities of the very technologies, that is, respectively their embedded “values” or their “politics”. However, this also needs to be understood more generally in terms of distributed governance in which the “values sensitive design” of the very data design of AI is only one component. Thus, when we examine the ethical implications of AI-based systems, we need to look at the way in which they are distributed among active human and non-human actors and we need to understand them as components of complex social, cultural and political environments. This also means that we cannot just design a moral norm into a technology and in this way produce a “moral machine”, we also need to address the way in which ethical implications evolve in environments of distributed moral agencies between human and non-human actors (Latour, 1992, 2002) that
are in constant negotiation with use, laws and standards, culture and the society. Social and ethical implications of AI systems are always the result of a combined network of actions and competences distributed between different agencies – the data design, the engineers, the users, norms, laws and culture etc. If we for example consider the gender bias of a “word embedding“ method for online search, it is not only a property of a particular AI design. The bias and discrimination result from a distribution of agency between the machine learning model (non-human actor) that is actively amplifying existing human bias in its training data, learning and evolving from Google news articles (society and culture) and personalized (the user data), prescribed by data design (the developer’s intentions), developed, accepted and enacted in society as an “objective” representation of information (culture), particular interpretations and implementations (or lack thereof) anti-discrimination or data protection legal frameworks (law), engineering standards and methods (scientific paradigms) and a range of other factors. This is why we need to think of a human-centric approach to the data interests of AI as not only a framework for the design of an AI technology, but moreover as a framework for the design of the “ethical governance” of AI in general.

Winfield and Jirotka (2018) present a case for “a more inclusive, transparent and agile form of governance for robotics and artificial intelligence (AI) in order to build and maintain public trust and to ensure that such systems are developed for the public benefit” (p. 1). Ethical governance, they argue, goes beyond just effective and good governance, but is “a set of processes, procedures, cultures and values designed to ensure the highest standards of behaviour” (Winfield & Jirotka, 2018, p. 2). They address the socio-technical nature of AI development and adoption in which the technical is intrinsically intertwined with the social and vice versa. Governing the development of robotics and AI with an ethical framework therefore requires a diverse set of approaches from those at the level of individual systems and application domains to those at an institutional level. Winfield and Jirotka are addressing the “ethical governance” instruments that companies need to develop and adopting AI ethically. But here I also want to propose that the “ethical governance” of AI concerns the applied components of AI ethics in the general governance of socio-technical systems with a distributed form of power in which data interests play a key role.
In what follows, I derive data interest themes from the “Ethics Guidelines for Trustworthy AI” (2019) developed by the AI HLEG appointed by the European Commission. It is important to note here that in isolation the guidelines do not ensure a human-centric AI development. That is; as stand-alone principles they are not the solution to an ethical problem. Winfield and Jirotka (2018) present a map that connects ethical principles with emerging standards and regulation, including a level of verification and validation. They also argue that “while there is no shortage of sound ethical principles in robotics and AI, there is little evidence that those principles have yet translated into practice, i.e. effective and transparent ethical governance.” (p. 9) That is; ethics guidelines, such as the AI HLEGs do not “govern”, but they may inspire, guide and even set in motion political, economic and educational processes that foster an ethical “design” of the big data age, which means everything from the introduction of new laws, the implementation of policies and practices in organisations and companies, development of new engineering standards, to awareness campaigns among citizens and educational initiatives.

**Ethics Guidelines for Trustworthy AI**
The High-Level Expert Group developing the guidelines consisted of 52 members representing stakeholder interests from industry, civil society and academia. Although set up in a policy context, the group was not a policy body per se. The European Commission sets up multi-stakeholder high-level expert groups to inform their policies in different areas. These are independent, and the Commission does not participate directly in the groups’ work, nor do they necessarily use the work of these groups directly in policymaking. The ethics guidelines were developed over a one-year period and included a public consultation process for the first draft in which comments were received and incorporated in the text. Although this very multi-stakeholder process by itself would be a relevant focus for a data interest analysis, I in this article focus only on the very text of the guidelines in order to illustrate a negotiation of data interests with a human-centric applied ethics approach.

The guidelines specifically aim to provide AI practitioners with a methodology for incorporating a human-centric framework in the design of AI and as follows also for the governance and management of the data of AI. Three core components of Trustworthy AI are proposed in the
guidelines. AI should be lawful—respect laws and regulations, ethical—respect ethical principles and values, and robust—from a technical perspective and with consideration of its social environment. These three components are the essential framework within which the concept of Trustworthy AI is interpreted. First of all, they refer to an existing legal framework; that is, the guidelines are not an alternative to or an interpretation of law. The first component, legal compliance, is therefore assumed as the basis of Trustworthy AI. The second “ethical” component is expanded in four ethical foundational principles: respect for human autonomy, prevention of harm, fairness and explicability. The third component is meant to ensure the robustness of the AI system to guarantee the safe, secure and reliable behaviour of the system to prevent unintended adverse impacts. The guidelines also present seven key requirements that AI systems should meet in order to be deemed trustworthy: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental accountability and auditability. An assessment list is here included, aimed at operationalizing these key requirements when developing, deploying or using AI systems. This list was revised in 2019–2020 in a piloting phase to include input from companies and institutions developing AI in different sectors.

The Five Data Interests and an Example

In the following, I illustrate a human-centric approach to the data interests of AI development based on an analysis of the data metaphors of the AI HLEG ethics guidelines. I suggest five clusters of themes and sets of questions that might help forward an exploration of data interests in AI design and development: data as resource, data as power, data as regulator, data as vision and data as risk. I then exemplify a human-centric approach to data interests in a case in which an AI design takes explicitly point of departure in the data interest of the individual human being.

This analysis is not intended as a comprehensive framework the implementation of a human-centric approach to data interests by design, nor does it represent a complete analysis of the data interests represented in the ethics guidelines. Importantly this is only a values-based interpretative framework that complements but does not by any means replace the implementation of legal frameworks. What I mean to do here is
to present a human-centric “compass” that may help guide an applied ethics by design approach to data interests in AI development and governance.

I. Data as Resource

*Who or what provides the data resource? Who or what has an interest in the data resource? How is the data resource distributed and how does the human being benefit?*

The “data as resource” cluster concerns the very distribution of data resources among involved interests in the data design. In the ethics guidelines, data is considered a resource, metaphorically separated from that which it represents (a person or an artifact). Data can be “provided”, “accessed”, “gathered”, “labelled”, “extracted”, “used”, “processed”, “collected”, “acquired” and “put” into a system in a “structured” or “unstructured” (words used throughout the guidelines to describe the data of AI). A resource is a corporeal and spatially delineated thing, something we can be in possession of, place in containers or create boundaries around, store and process, and it is something we can be with or without. The resource metaphor is common in public discourse on big data (Puschmann & Burgess, 2014). Sarah M. Watson (n.d.) refers to the dominant metaphors for personal data in public discourse as “industrial”, as if it were a “natural resource” to be handled by “large-scale industrial processes”. Mayer-Schonberger & Cukier (2013) depict data as the raw material of a big data evolution of the industrial age. But the “data as resource” metaphor actually denotes two different things. One type of resource in society is indeed the raw material of industrial processes that is processed and turned into products. Another type of resource is the kind that makes us stronger as individuals. Being resourceful also means that you as an individual have the capacity, and physical, psychological and social means. The first type of resource mentioned here is tangible and material, the other being social and psychological. But both are what Lakoff and Johnsson (1980, p. 25) would refer to as “container” metaphors, entities with boundaries that we can handle and reason about. Subsequently, data as a resource can be protected and governed in very tangible ways, and in each case micro and macro stakeholder interests in these governance and resource handling frameworks are involved. Very broadly speaking, industries have an interest in the raw material of their data-based business models: political players have an interest in providing rich data infrastructures for AI.
innovation to compete in a global market; the engineer has an interest in volumes of data to train and improve the AI system; individuals have an interest in protecting their personal data resourcefulness or even enhancing their data resources by creating their own data repositories and accordingly benefitting directly from these (as the “data trust” or “personal data store movement” represents). Obviously, treating the psychological and social resources of individuals as material resources in an industrial production line represents a conflict of data interests. But, in addition, informational asymmetries also create very tangible social and economic gaps between the data rich and the data poor, which is a conflict of interest on a more general structural level of society. In the guidelines, there is a reference to data as the raw material resource of AI technologies (“its evolution over the last several years has been facilitated by the availability of enormous amounts of digital data” p. 35). However, data and the governance of this is first and foremost associated with privacy and the protection of the individual human being (the third of the guidelines’ requirements, p. 17). Closely connected with the ethical principle “prevention of harm”, the protection of data as a social and psychological resource of human beings is in the guidelines described as “protection of human dignity as well as mental and physical integrity” (p. 12). This includes particular attention to socially vulnerable groups or individuals. Proper governance of the personal data resource is described as the primary means to “allow individuals to trust the data gathering process” (p. 17) and consequently forms a foundation the development of “Trustworthy AI”. To provide an example, an AI developer will make use of different resources of data for the design of an AI system. These data resources they have to either process on their own computers or they can use the more powerful cloud-based AI platforms of for example Google or Amazon. However, by doing this they also have to share their data resources with these actors. If prioritizing individuals’ data interests, when dealing with for example personal data, they have to trust that these companies do the same. This entails a data interest design choice. Moreover, different actors may have interests in the data resource. For example, if the developer is creating an AI system for assessing the performance of employees, the employee might have an interest in creating a bigger data resource that can assess minute details of the employees work day. The employer might even want to collect data from outside the work place from the employees’ social
media presence for the AI system to predict potential risks to the workplace, such as for example job searches. The employees on the other hand have an interest in keeping the data resource to a specific limit and a union representing these employees might also want to get involved to safeguard the data interest of the people they represent.

II. Data as Power

*Who is empowered or disempowered by the data access and processing? Does the data design support the human being’s power and data agency? How are conflicts between different data interests resolved to the benefit of the human being?*

The second cluster, “data as power”, is closely connected with the first cluster, “data as resource”, as the distribution of resources also constitutes a distribution of power. A society is based on balances of power between social groups, states, companies and citizens. A democratic society, for example, represents one type of power structure in which the governing powers are always balanced against the level of citizen power. Distribution of data/information amounts to the distribution of power in society. David Lyon envisions an “ethics of Big Data practices” (2014, p. 10) to renegotiate an unequal distribution of power embedded in technological big data systems. In the guidelines, changing power dynamics are addressed in light of the information asymmetry between individuals and the institutions and companies that collect and process data in digital networks, where AI systems are the interlocutor (the agents) of this type of informational power: “Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens.” (p. 12). Access to or possession of data is here associated with dominance of some societal groups and or institutions, businesses over others, but also with the very function of democratic and or “fair” processes, where access to information and explanation of data processing (“the principle of explicability”, p.13) is the basis of “accountability” and “fairness” (“Without such information, a decision as such cannot be duly contested.”, p. 13). Data design solutions suggested are here associated with the concept of “human agency” that supports “informed autonomous decisions regarding AI systems” or “informed choices in accordance with their goals” (p. 16).
For example, access to the data and data processes of a system can provide an investigative journalist or researcher with the power to challenge an AI system. The researcher has an interest in the data of the system that can be inhibited by a company’s interest in keeping algorithms and data design proprietary. In connection with the Machine Bias study of the Compass software, the researchers did not have access to the calculations used for defendants risk scores. At one point they received the basics of the future-crime formula from the company behind, Northpointe, but it never shared the specific calculations, which it said are proprietary (Angwin et al, 2016).

III. Data as Regulator

Which interests will the implementation of law serve and prioritize? Does the data design of the AI system truly enhance the values of the law, or does it just barely comply? Are there conflicts between big data interests and the embedded values to protection of the individual human being of a legal framework such as the GDPR? How are they resolved in the design?

The third cluster, “data as regulator”, represents the legal enforcement of the power balancing of data interests. In an ideal constellation, law and technology design supplement each other. Technology design is a type of “regulator” that either protects or inhibits legal values. “Code is law”, as Lawrence Lessig (2006) formulates it quoting Joel R. Reidenberg’s concept “Lex Informatica” that links technological design choices and rule-making with governance in general and understanding the two as inseparable, with the first enforcing the other (Reidenberg, 1997, p. 555). Although it is emphasized that the guidelines do not replace existing laws, and therefore the first component of Trustworthy AI, “lawful AI”, is only a reiteration of compliance with legal frameworks, it can be argued that the guidelines still treat the “Lex Informatica” of lawful AI: “The guidelines do not explicitly deal with the first component of Trustworthy AI (lawful AI), but instead aim to offer guidance on fostering and securing the second and third components (ethical and robust AI)”(p. 6). Throughout the text, the design of AI is associated with the implementation of legal principles. The data design of an AI technology is here described as essential to the auditability and accountability of an AI system (p. 19) that ensures that it can be assessed and evaluated as well as
be able to implement the principle of explicability (“Without such information, a decision cannot be duly contested.” p. 13). In this way, the “data as regulator” metaphor emphasizes the role of a particular data design in the legal implementation and realization of law in society (“AI systems can equally enable and hamper fundamental rights.” p. 15). That is also to say that even though compliance with the European General Data Protection Regulation (GDPR) is considered in the development of an AI technology, it may or may not actually realize the legal principles of this law in its very data design. In a micro-engineering context, the design challenges of a data-intensive technology such as AI are, for example, the development of a data design implementing principles such as data minimization, privacy (data protection) by design or the right to (information) an explanation. Different stakeholder interests in the data design may also be in conflict with legal frameworks such as the Fundamental Rights Framework in which the data interests of individuals is a primary value. For example, non-democratic state actors have an interest in data for social control purposes; business interests in tracking and collecting data to enhance their data-based business models without consideration of individuals’ rights, or similarly scientific interests in improving research with big data analytics without these considerations. The ethics guidelines mention some of the greatest legal challenges of AI in a European fundamental rights framework—mass surveillance, discrimination, the undermining of democratic processes—and propose alternative design options embodied in the concept of Trustworthy AI.

IV. Data as Eyes

*Can the human beings involved in the design and implementation “see” the data processes and their implications? What does the data design see (the training data) and then perceive (how is it instructed to act on training data)?*

The fourth cluster, “data as eyes”, represents the very agency in the data design of data interests that here may be equated with vision (what we are able to see or not, and how we see it) in digital data systems. In an ideal constellation, vision is an effortless extension of our eyes. However, in a digital data-based environment, the instruments (our digital eyes) that we use to see and perceive our environment are quite literally extended into data design that constitutes a management of what we can see.
As previously discussed, data design also functions as a moral agent in that it prescribes and manages our active engagement with the information it handles as well as the digital information infrastructure in which it exists. In the ethics guidelines, eyes and vision are embodied in data systems. While data is described as the technology's sensory system on which it develops its mode of action ("perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal", p. 36), data is also the "eyes" of the individual. Data will for example "yield the AI system's decision, including those of data gathering and data labelling as well as the algorithms used..." (p. 18). The concept of "transparency" is used in the most literal sense to denote our ability to see the data processes that should be documented and made traceable. "Black box" algorithms' (Pasquale, 2015) processing of data may in this context blur our vision or make us blind to the reasoning of an AI system. The management of visibility, that is, the very architecture of visibility of emerging technological environments can be said to constitute a mode of social organization and distribution of powers (Brighenti, 2010, Flyverbom, 2019). What is made visible, what remains invisible and importantly who is empowered to see through the social organization of visibilities directly influences the agency of interests in society. The eyes as the data design of an AI technology do exactly that. To provide an example, we might create an AI system to analyze the data of people on social welfare benefits. This may have a dashboard for the public institution's social workers, which provides general data statistics, fraud detection and risk scores on individuals. The dashboard is our eyes within the AI system. In this case, only the social worker’s data interest has eyes and so does the public institution's interest in controlling public resources and optimizing work processes. But we could also think of other types of data design in which the people on social welfare have eyes through data access and hence agency to, for example, add and correct flawed data or personalize the services provided to them.

V. Data as Risk

*To whom or to what could the data design be a risk? Who or what has an interest*
in preventing and managing the identified risks? How are conflicts and/or alignments between identified risks to the human being and interests in managing the risks resolved to protect the human being from risks?

In the fifth and last cluster, “data as risk”, data interests are correlated with the act of assessing the risks of data design. The risks of the development and deployment of AI technologies are addressed generally in the guidelines as the potential negative impacts of AI systems on democracy and civil rights, on the economy, and on markets and the environment. The risks are the harms of AI that should be anticipated, prevented and managed (p. 2). In particular, in the third core component (“robust AI”), a preventative risk-based approach to AI is emphasized: “Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner, such that they reliably behave as intended while minimising unintentional and unexpected harm” (p. 16). Risks are not just a particular concern of these guidelines, but of contemporary politics, business conduct and public discourse in general. Ulrich Beck described this societal preoccupation with risk prevention and management in his seminal book The Risk Society (1993) as an uncertainty produced in the industrial society, and as the result of a modernization process, in which unpredictable outcomes are emerging and accumulating (Beck, 1992). “Risks are not ‘real’, they are ‘becoming real’”, Beck (2014, p. 81) later proclaimed when describing a development of concern with “world risks”. They take the shape of “the anticipation of catastrophe” (Beck, 2014, p. 81) and their management as an “anticipation of further attacks, inflation, new markets, wars or the restriction of civil liberties” (p. 81). Importantly, the depiction of a risk “...presupposes human decisions, human made futures (probability, technology, modernization)...” (Beck, 2014, p. 81)

In the AI HLEG guidelines, the data of AI is described as laden with potential risks that should be prevented and managed. As described in the four other metaphorical clusters, data is generally associated with the management of risks to resources in society, to democracy, to the rule of law and to agency through visibility. But in this last metaphorical cluster of the guidelines, data is a risk per se that must be anticipated and managed. Criminals can for example “attack” the data of an AI system and “poison” it (p. 16), data can “leak”, data can be “corrupted by malicious intention or by
exposure to unexpected situations” (p 16), and it can pose a risk to the environment (“Did you establish mechanisms to measure the environmental impact of the AI system’s development, deployment and use (for example, the type of energy used by the data centres)?”, p. 30). Data by itself also has risky properties; for example, words such as “malicious” or “adversarial” are also used in the guidelines to describe data. If we continue with the notion that risks are not “real” per se but based on our own predictions about possible future scenarios, then we may also assume that their proposed management and prevention is the product of interests and motives. On the AI development side, an AI engineer training an AI technology will have an interest in the risks posed to the quality and accuracy of the training data, while a data protection officer will consider risks posed to identified individuals with a data protection impact assessment. On the AI deployment and adoption side, an individual will also have a “data-as-risk” interest in keeping their personal data safe from unauthorized access, while an anti-terror intelligence officer’s risk scenario involves the detection of terrorist activities and might therefore consider the end-to-end encryption of a data design a risky design choice. The consideration of different “data-as-risk” scenarios guided by the different interests involved in these will direct disparate design choices that might be aligned but might also be in conflict.

**An Example: P7006 The Personal Data AI Agent**

The IEEE P7006 standard for Personal Data Artificial Intelligent (AI) Agents (PDAIAs) which is one of the IEEEs P7000s series of ethically aligned design standards for autonomous and intelligent systems. The standard was work in progress at the time of writing this article, but the PAR and work in progress (wp) principles (P7006, 2020) can be used to illustrate how reflections on the operationalization of some of the thematic clusters of a human-centric framework for data interests can be translated into a technical standard for the design of an AI technology.

Personal AI assistants are software programs that interact with individuals answering questions and performing different tasks for them. They are based on various AI components such as voice recognition and machine learning to act on behalf of and respond to individual users in a personalized manner. Personal AI assistants have been popularized by leading commercial actors as integrated components of mobiles, tablets or
speakers with for example Amazon’s Alexa, Apple’s Siri, Google Now and Microsoft’s Cortana (Bonneau, Probst & Lefebvre, 2018). As virtual agents integrated in individuals’ private sphere at home and out that recognize voices, store and process users’ personal data, several critiques and concerns have raised in the public debate. In particular privacy concerns and the prioritization of the power and interests of the commercial actors and other actors versus that of individuals (Lynsky, 2019, Nelius, 2020, Chung et al., 2017, Maedche et al., 2019).

Against this background, the P7006 standard’s scope is explicitly defined in its Project Authorization Request (PAR) as one that prioritizes the data interest of the individual human being: "This standard describes the technical elements required to create and grant access to a personalized Artificial Intelligence (AI) that will comprise inputs, learning, ethics, rules and values controlled by individuals.” (P7006, 2017) It is furthermore in the PAR envisioned as an agent that first and foremost manages the data interests of individuals enabling them to “safely organize and share their personal information at a machine-readable level and enable a personalized AI to act as a proxy for machine-to-machine decisions (P7006, 2017). The PDAIA is trained and evolves with the data of an individual. As such the main data resource is that of the individual. The data interests in this data are described in the scope and purpose of the standard: “The primary stakeholders for this standard are the individuals whose information is utilized for any data interaction, but also includes the academics, engineers, programmers, marketers or technologists of any kind wishing to utilize said data” (P7006, 2017). However, the data interest benefitted most profoundly by the data resources is that of the individual in that the AI data design is meant to serve as an agent for the individual that “negotiates their individual rights and agency in a system of shared social norms, ethics and human rights that also foresee and helps the individual mitigate ethical implications of data processing” (P7006, 2017). This prioritization of the individual’s interest in data is sought operationalized throughout the wp principles, for example in one principle that is meant to ensure that the data resource is maintained by the individual human. The interest of the individual human being is moreover prioritized in the power distribution of the data design. For example, the same principle aims to ensure that the power over databased decisions remains with the individual when stating that the PDAIA will remain semi-
autonomous and never fully independent of the individual while making decisions. As such, another principle concerns “Interruptible & Retractable” decision-making and among others includes the incorporation of a “Kill Switch” at root-level design that the individual user can use to halt and or reverse the actions of an agent. The data design’s management of visibility is moreover specifically meant to provide the individual human being with vision of the data processes and thereby agency. The first principle for example entails “human-interpretable ”transparency”” which is further delineated with the provision: ““Black box” processes must be able to be tracked in a way that creates a human-readable chain of custody for all decisions”. The vision of the individual human being is also extended to include information regarding third-parties that have influenced a decision. Data as a type of regulator is considered with specific consideration of “privacy” design rather than data protection and security only for example in one principle that considers a “Privacy-oriented dynamic UI/UX. A user can easily toggle Agent privacy settings to enable updating system settings to facilitate differential privacy depending on the occasion.” Finally, data is primarily considered a risk to the human being and accordingly one principle considers the safety of the user based on user-centered agency, protection against ”physical, emotional or cognitive harm” and safeguards from revealing compromising information from law enforcement.

Conclusion

Over a very short period of time there has been an acceleration in the integration of data-intensive technologies such as AI —a complex and advanced machinery designed to act on data, predict behaviour and trajectories, and point lives, economies, politics, society in specific directions— in all areas of our lives. AI technologies constitute a distribution of informational resources in society, and they therefore have a direct influence on the balance of agency among societal actors.

In this article, I have made a case for a data interest analysis of AI technologies to identify data interests in their data design and to explore how these interests are empowered or disempowered by design. I have also put forward a specific applied ethics approach to data interests based on the human-centric ethical governance framework presented in the EU High-Level Group on AI’s Ethics Guidelines for Trustworthy AI.
AI systems are not impartial or objective representations of the world. They reinforce societal power relations and are invested with the interests of the powerful, and they therefore trigger actions with ethical and social implications. Big data industries have an interest in harvesting and accumulating data, improving their machine-learning models or earning a profit on personalized marketing. Public institutions have an interest in streamlining and controlling public administration. Governments have an interest in controlling, viewing and predicting risks to their nations from data. Global industry and intergovernmental bodies have an interest in gaining a competitive advantage with data. Obviously, interests are not as uniform as this, nor are they simple determinants of action in society. But as we see time and again when ethical implications of AI systems materialize into social implications, data interests are present in the design of AI and are provided with agency or no agency. The result is increasing informational asymmetries among people, such as between civil servants, developers or people from minority groups and societal stakeholders, such as industries, governments and citizens.

An applied ethics approach to AI systems highlights the social power relations and interests at play in the development and societal adoption of AI. Narrowing down an approach as such to the data interests in AI allows us to consider the built-in trade-offs between these different interests, and it opens up the possibility of an exploration of alternative data designs that may solve conflicts of interests in a way that benefits the human being. Here, fostering an awareness of the primacy of the human interest over commercial, state, and even scientific interests in the development of big data AI systems is one such awareness that needs a strong position in the AI ethics debate.

---

1 To investigate the themes, I used a metaphorical analysis (Lakoff & Johnsson, 1980). I first went through the text to detect metaphors concerning data connecting them with common metaphors concerning data in public discourse. Then I organized the text according to the core metaphors I found with a focus on the data interests they represented, their conflicts and finally how they were sought balanced in a human-centric framework.

2 I am the vice chair of the standard working group which is chaired by Ken Wallace. The PAR and work in progress principles are publicly available on the P7006 working group website. The work in progress draft principles are created by the members of the P7006 working group with the objective to guide the more detailed work on the standard. They are not complete, subject to negotiation and revision, and should therefore also be viewed as such. The specific draft principles that I refer to here were substantially revised.


Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T., 2016 “Man is to computer programmer as woman is to homemaker? Debiasing word embeddings” 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain


on AI, ethics, and society—AIES ’18 (pp. 60–66). ACM Press.
https://doi.org/10.1145/3278721.3278745

Deleuze, G., 1992 "Postscript on the societies of control" October, 59, p. 3-7.


Friedman, B., Kahn, P. H., Jr., & Borning, A., 2006 “Value sensitive design and information systems”. In P. Zhang and D. Galletta (Eds.), Human-computer interaction in management information systems: Foundations (pp. 348–372). M.E. Sharpe.


GDPR/Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of
such data, and repealing Directive 95/46/EC (General Data Protection Reguation)


Hasselbalch, Gry (forthcoming): Culture by Design – A Data Interest Analysis of the European Policy Agenda, FirstMonday

Hughes, T. P., 1987 “The evolution of large technological systems”. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), The social construction of technological systems (pp. 51–82). MIT Press.


P7006, June 2020 “Work in Progress Principles for P7006” https://sagroups.ieee.org/7006


Lynsky, D., 2019, 9th Ocetober, in press “‘Alexa, are you invading my privacy?’ – the dark side of our voice assistants”, The Guardian.

Mayer-Schönberger, V., & Cukier, K., 2013 Big data: A revolution that will transform how we live, work and think. John Murray.

Maedche, A., Legner, C., Benlian, A. et al., 2019 "AI-Based Digital Assistants". Bus Inf Syst Eng 61, 535–544

Nelius, J., 2020, September 4, in press "Amazon’s Alexa for Landlords
Is a Privacy Nightmare Waiting to Happen”, Gizmodo.


OECD, 2019 “Principles on Artificial Intelligence” in OECD Council Recommendation on Artificial Intelligence, OECD/LEGAL/0449, 22.05.2019


OECD, 2019 “Principles on Artificial Intelligence” in OECD Council Recommendation on Artificial Intelligence, OECD/LEGAL/0449, 22.05.2019


