

ÖAW

AUSTRIAN
ACADEMY OF
SCIENCES



NARRATIVES OF DIGITAL ETHICS

**AGIDE (ACADEMIES FOR GLOBAL
INNOVATION AND DIGITAL ETHICS) REPORT**

IMPRINT

PUBLISHER

Austrian Academy of Sciences
Dr. Ignaz Seipel-Platz 2
1010 Vienna
<https://www.oeaw.ac.at/en>

EDITOR OF CONTENT

Univ.-Prof. Dr. iur. Christiane Wendehorst, LL.M. (Cantab.)

COVER PHOTO

© Shutterstock

EDITORIAL OFFICE

Jana Winter, BA

COPY EDITING

Nicola Wood

GRAPHIC DESIGN EDITING

Mag. Daniela Seiler

PRINT

Prime Rate, Budapest

All rights reserved. Copyright © 2024. The responsibility for the content and copyright for the individual contributions lies with the respective authors. Opinions expressed represent the views of the respective authors and do not necessarily reflect those of the OeAW.

<https://doi.org/10.1553/978OEAW97058>

NARRATIVES OF DIGITAL ETHICS

**AGIDE (ACADEMIES FOR GLOBAL INNOVATION
AND DIGITAL ETHICS) REPORT**

ACKNOWLEDGEMENTS

The research project “Academies for Global Innovation and Digital Ethics” (AGIDE) led by the Austrian Academy of Sciences (OeAW) seeks to explore the various (cultural) dimensions of digital ethics and global innovation – that is, how technologies are transforming our societies and global structures – by taking off the “Eurocentric glasses”, embracing differences, and going beyond familiar stereotypes or clichés. To this end, the Austrian Academy of Sciences has been cooperating with ten other Academies of Sciences from all over the world. The OeAW would like to take this opportunity to thank the following partner academies for their excellent and fruitful cooperation in the AGIDE project (in alphabetical order):

- The Australian Academy of Science (AAS)
- The Academy of Science of South Africa (ASSAf)
- The Brazilian Academy of Sciences (ABC)
- The Chinese Academy of Sciences (CAS)
- The German National Academy of Sciences Leopoldina
- The Israel Academy of Sciences and Humanities (IASH)
- The National Academy of Sciences (NAS)
- The Pakistan Academy of Sciences (PAS)
- The Polish Academy of Sciences (PAN)
- The Royal Netherlands Academy of Arts and Sciences (KNAW)

List of all Members of the AGIDE Working Group (in alphabetical order):

- Teki AKUETTEH
- Virgilio ALMEIDA (ABC)
- Nisreen AMEEN
- Payal ARORA
- Markus ASPELMEYER
- Anat BEN-DAVID
- Keith BROWN (NAS)
- Yik Chan CHIN
- Magdalena EITENBERGER
- Max HALLER
- Soraj HONGLADAROM
- Dariusz JEMIELNIAK (PAN)
- Te Taka KEEGAN
- Katrina LIGETT (IASH)
- Astrid MAGER
- Yishu MAO
- Keymanthri MOODLEY (ASSAf)
- Herta NAGL-DOCEKAL
- Sarayu NATARAJAN
- Ruben George OLIVEN (ABC)
- Manisha PATHAK-SHELAT
- Barbara PRAINSACK
- Emma RUTTKAMP-BLOEM
- Albrecht SCHMIDT (LEOPOLDINA)
- Zabta Khan SHINWARI (PAS)
- Shoko SUZUKI
- Linnet TAYLOR (KNAW)
- Toby WALSH (AAS)
- Astrid WEISS
- Christiane WENDEHORST
- Jana WINTER
- Yi ZENG (CAS)

The short biographies of all members can be found at the end of the report (“Annex”).

Members of the OeAW Steering Committee who have advised on the AGIDE project (in alphabetical order): Markus ASPELMEYER, Max HALLER, Astrid MAGER, Herta NAGL-DOCEKAL, Barbara PRAINSACK, Astrid WEISS, Christiane WENDEHORST.

Primary Authors: Christiane WENDEHORST, Magdalena EITENBERGER, Jana WINTER, Astrid MAGER, Barbara PRAINSACK, Astrid WEISS.

Further Authors (in alphabetical order): Teki AKUETTEH, Virgilio ALMEIDA (ABC), Nisreen AMEEN, Payal ARORA, Markus ASPELMEYER, Anat BEN-DAVID, Keith BROWN (NAS), Yik Chan CHIN, Max HALLER, Soraj HONGLADAROM, Dariusz JEMIELNIAK (PAN), Te Taka KEEGAN, Katrina LIGETT (IASH), Yishu MAO, Keymanthri MOODLEY (ASSAf), Herta NAGL-DOCEKAL, Sarayu NATARAJAN, Ruben George OLIVEN (ABC), Manisha PATHAK-SHELAT, Emma RUTTKAMP-BLOEM, Albrecht SCHMIDT (LEOPOLDINA), Zabta Khan SHINWARI (PAS), Shoko SUZUKI, Linnet TAYLOR (KNAW), Toby WALSH (AAS), Yi ZENG (CAS)

TABLE OF CONTENTS

EXECUTIVE SUMMARY	5
THE AGIDE PROJECT – INTRODUCTION	7
ON DIGITAL ETHICS	8
ON THIS REPORT	9
PART A: DIGITAL ETHICS – WITH A FOCUS ON DIFFERENCES	10
UNIVERSALIST APPROACHES TO DIGITAL ETHICS	10
Existing Sets of Principles	10
DIFFERENCES BENEATH THE SURFACE	12
Local Contexts and Applications of Digital Ethics Principles	13
Why Situatedness Matters	14
PART B: THE AGIDE PROJECT – EMPIRICAL FINDINGS	16
THE EMPIRICAL ANALYSES – INTRODUCTION	16
THE AGIDE PROJECT – GENERAL METHODOLOGY	16
THE AGIDE PROJECT – LIMITATIONS	17
THE AGIDE INTERVIEWS: ANALYSIS	18
The AGIDE Interviews – Introduction	18
Meaningful Differences in Digital Ethics	18
Relevant Technologies: Which Current Developments Impact Digital Ethics?	20
Digital Ethics: Mediated by Ownership, Governance and Power?	21
Digital Exclusion and Marginalisation	24
Technological Solutions for Digital Ethics Problems	27
Key Areas of Applied Digital Ethics	27
Digital Ethics and Digital Futures: Current Realities, Future Imaginaries	29
THE AGIDE WORKSHOPS: ANALYSIS	30
Introduction	30
Exploring A Good Digital Future For All	30
The AGIDE Workshops I, II & III: Visions For A Good Digital Future	31
What the Global North Has Been Overlooking	31
Interconnectedness, Ecology and Harmony	32
The Future of Work, Youth Empowerment and Finding a Balance	32
Democracy, Human Rights and the Welfare State	33
Move Fast and Break Things: The “First To Market” Approach to Innovation	34
A Brief Word of Conclusion	34
THE AGIDE USE CASE SCENARIOS: ANALYSIS	34
The “Use Case” Scenarios – Introduction	34
Public Space	35
Social Media	36
AI Literacy	36
Non-human Agency	37
Data Communities	37
Climate	38
PART C: MAPPING DIGITAL ETHICS – THE NARRATIVES	39
SYNTHESISING THE EMPIRICAL FINDINGS	39
MAPPING DIGITAL ETHICS – THE NARRATIVES	39
AN INITIAL MATRIX FOR ETHICAL NARRATIVES	40

EIGHT KEY DIMENSIONS OF DIGITAL ETHICS NARRATIVES	41
Fundamental Ethical Assumptions	41
The “Protagonist” and Its Role	41
The “Conflict” and the “Antagonist”	42
The Wider “Plot” (What Should Happen)	42
CHARACTERISTIC NARRATIVES OF DIGITAL ETHICS – THE MACROPERSPECTIVE	43
“Coloniality”-Type Narrative Patterns	43
“Beneficiary”-Type Narrative Patterns	44
“Harmony-Opportunity”-Type Narrative Patterns	45
“Silicon Valley”-Type Narrative Patterns	46
“GDPR”-Type Narrative Patterns	48
CHARACTERISTIC NARRATIVE PATTERNS – MOVING TO THE MICROPERSPECTIVE	49
Differences within a Region – the Example of the European Union	49
Differences within a Population – the Example of India	53
PART D: IMPLICATIONS AND OUTLOOK	56
THE POWER OF ETHICAL NARRATIVES	56
IMPLICATIONS FOR FURTHER RESEARCH AND POLICY-MAKING	58
CLOSING REMARKS	59
REFERENCES	60
LIST OF FIGURES	65
ANNEX: MEMBERS OF THE AGIDE WORKING PARTY	66

EXECUTIVE SUMMARY

The digital transformation has brought about an unprecedented degree of global interconnectedness, accompanied by increasing efforts to formulate universal ethical guidelines for dealing with emerging digital technologies. The relative ease with which countries around the world seem to agree on universal action-guiding principles of digital ethics along the lines of “fairness”, “transparency” and “accountability” seems to contrast sharply with the vast differences in technology adoption that we see around the world, and also the vast differences in attitudes towards technology. These differences can influence the salience of ethical concerns in different settings. The project AGIDE, which stands for “Academies for Global Innovation and Digital Ethics”, seeks to embrace this diversity of perspectives on digital ethics, exploring where differences lie and how these differences might be conceptualised beyond existing stereotypes.

In order to get a clearer picture of the key issues, similarities and differences, the starting point was over 75 qualitative interviews with expert voices from around the world, as well as deep dives in the course of three workshops held in April, June and October 2023. Taking a “situated” approach and considering local knowledge contexts, the results were then analysed and synthesised by an international working group comprising representatives from 11 academies of sciences on six different continents, as well as other eminent experts from around the world.

The AGIDE project found that there is a remarkable consistency in core values (such as “justice”, “dignity” or “privacy”) across different regions of the world. The fact that agreement has been reached on common principles and guidelines, such as the *UNESCO Recommendations on the Ethics of Artificial Intelligence* or the *OECD Principles on Artificial Intelligence*, is arguably a result of this fundamental agreement on shared core values. Contrary to initial expectations that the differences in perception and governance of digital opportunities and risks explored by AGIDE might result from discernible differences in emphasis on particular values, the data collected did not support such distinctions.

AGIDE’s research showed that major differences lie in something else: in the narratives of digital ethics. Narratives

are stories that are told repeatedly, consisting of a series of events that are selected and arranged in a particular order, often including central characters (protagonists, antagonists), a conflict and a plot. When narratives become dominant in a particular social setting, because they are shared by a wider group and/or promoted by influential actors, they can become powerful drivers of collective behaviour, and they shape how core values are operationalised and put into practice. Findings from the AGIDE project suggest that although core values are widely shared, digital ethics discourses in different world regions are dominated by different narratives about how these values are challenged, or how they can be protected, and why and how these values matter in the first place.

The data collected in the AGIDE framework was fitted into a matrix according to criteria that emerged from the interviews and workshops. This matrix attempts to categorise the various views and concerns expressed in relation to eight salient aspects, including underpinning ethical approaches, the primary point of reference (for example, the individual or the community) and its position vis-à-vis technology (for example whether it is primarily perceived as a victim, a beneficiary or an actor), the primary focus of ethical concern, and appropriate tools of governance. Taken together, the very particular approaches to these eight aspects result in characteristic patterns of storytelling that reflect distinct digital ethics narratives.

From a macro perspective, several characteristic narratives emerged, including what the report calls the “Coloniality”-type narrative, the “Beneficiary”-type narrative, the “Harmony-Opportunity”-type narrative, the “Silicon Valley”-type narrative, and the “GDPR”-type narrative (with GDPR standing for the EU’s General Data Protection Regulation). At the level of practice, these patterns of storytelling may shade into one another. Exploring the narratives from a micro perspective – looking at specific regions or populations – revealed a much greater diversity of patterns, highlighting the nuanced nature of these perspectives. There is no single answer as to what produces the differences in narratives. The data suggest that there is a range of different factors, and that the intersectionality of these factors influences the choice of approaches.

It was particularly insightful to observe that many patterns of storytelling are fluid to some extent, and that there are remarkable shifts as far as dominant narratives in a country or region are concerned. However, some narratives seemed to be very deeply entrenched. The European Union (EU) was presented as an illustrative example where the “GDPR”-type narrative seems embedded to an extent that might potentially prevent EU policymakers from effectively bringing about change.

Given the scope and timeframe of AGIDE, the findings highlight the need for further research to explore whether the narratives are causes or symptoms of the differences we perceive, or both. Further research is also needed on the factors that contribute to the development of specific narratives, both at the macro and at the micro level. Finally, we need to understand better the conditions that are conducive to the transformation of established narratives or that cause established narratives to resist even major shifts in the policies pursued, potentially hindering important policy changes.

AGIDE therefore hopes to contribute to a new global discourse and to policymaking that is attentive to differences across and within regions regarding digital ethics narratives and, importantly, is alert to the significance of resources and power. Accepting and embracing differences means allowing for different interpretations and implementations of shared values and being open to the possibility of the development of situated ethical codes, without forcing uniformity. Respecting equity, fairness and empowerment also means giving priority to enabling the conditions and possibilities for local implementation, including the development of genuinely local technologies, structures and solutions. Finally, understanding the deep connection between narratives, ethical principles and power negotiations also requires that ethical principles can be translated meaningfully into practice, backed up, where necessary, with policies and legislation.

THE AGIDE PROJECT – INTRODUCTION

The “digital” – as in the digital age, the digital space or the digital world – has become a shorthand not just for (novel) electronic and computerised technology, but also for the seemingly all-encompassing nature and large-scale transformations that have attended their normalised use and presence in the lives of people around the globe. With their potential to (re)define fundamental aspects of life and society, digital technologies also “require us to rethink the very basis of the ways in which we relate to one another”, as Woolgar (2002, p.1) urged over twenty years ago. The inherent tension of a digitalised world therefore lies in its potential to change, to innovate and to transform while also referring us back to the most fundamental aspects of human (and non-human) relations. At its core, such (re)thinking entails reflection on the underlying ethics – the moral values and different notions of “the good” – that ground and guide digital technology development, its use and downstream consequences. At the same time, the transformative potential of electronic information and communications technologies (ICTs) and the like charges any engagement with the digital or “digital ethics” with careful consideration of the future(s) expected, cared for and enacted, as well as those that are ignored and neglected in the present (Brown and Michael, 2003; Levitas, 2013; Jasanoff and Kim, 2015; Prainsack, 2022). Adam and Grove (2007, p. 108), discussing such “future matters”, pointedly summarise the connection of ICTs, futures and values as follows: “With networked ICTs that operate in a temporal context of both instantaneity and simultaneity, traditional relations and approaches to the future are unsettled.”

Such questions and the dual interest in digital ethics as a means to discuss and investigate digital futures lie at the heart of AGIDE and the following report. More concretely, *AGIDE is an exploration of the vast differences in relationships with and attitudes towards ICTs, as expressed in terms of digital ethics. It set out to engage with the diversity of perspectives on digital ethics, exploring where differences lie and how these differences might be conceptualised beyond existing stereotypes and hegemonic tropes.* To do so, AGIDE collected a vast amount of data between 2022 and 2024, originating with a team at the Austrian Academy of Sciences in Vienna and supported by academics and “digital experts” around the world.

AGIDE’s approach thus builds on scattered attempts to address current issues in intercultural digital ethics (e.g. Aggarwal and Floridi, 2020; Berberich, Nishida and Suzuki, 2020; Capurro, 2008), such as, for example, ethical pluralism (e.g. Ess, 2006 & 2020) and cosmopolitan ethics (e.g. Stan van Hooft, 2009). Also, it complements other lines of research that focus on differences, such as the study of moral scenarios in different cultural contexts and countries (e.g. the *Moral Machine Experiment* by Awad *et al.*, 2018), the study of “global AI narratives” (Cave, Kanta & Dillon, 2020) and (future) imaginaries (Jasanoff and Kim, 2015; Mager and Katzenbach, 2021), as well as efforts to investigate an “anthropology of morality” (Heintz, 2018 & 2021) and ethnographic narrative practices (*Introduction: The Nexus of Anthropology and Narrative: Ethnographic Encounters with Storytelling Practices*; Götsch and Palmberger, 2022). Nevertheless, AGIDE also draws on related work, which does not exclude the possibility that differences may result from factors that cannot be adequately captured by “culture” in the traditional sense, such as socio-political narrative analysis (e.g. Wodak, Reisigl, and de Cillia, 2022; Bradford, 2012 & 2023; Pagallo, 2024) and narrative and technology ethics (e.g. Reijers and Coeckelbergh, 2020).

This report therefore presents a theoretical basis for the generative engagement with globalised ethics – often articulated as ethical principles such as “fairness”, “transparency”, “accountability”, “preservation”, “continuity” and “conservation” – alongside its localised (i.e. “situated”) expression and enactment in context (Part A). This framework served as guidance for the empirical data collection, which involved 75 semi-structured qualitative interviews with experts in a total of 28 countries and two input workshops to glean key topics for digital ethics around the world, as well as a group-based discussion of potential digital future(s) scenarios to better understand visions of a “good digital future” (Part B). Together, this data was systematically grouped and mapped out along eight dimensions representing key reference points for crafting situated arguments around digital ethics in a matrix. The use of such storytelling or “ethical narratives”, as we will call them, applies the flexibility of connecting plot points, actors and storylines within a shared framework as a tool

to project forward and imagine possible futures (Part C). Finally, we summarise the possibilities entailed in a shared value framework that can work flexibly to inform policy, regulations and practices to suit particular contexts. AGIDE employed an expert-focused bottom-up approach, exposing various points of departure for future research, especially on the question of how shared norms are articulated and operationalised differently across sectors, population groups and contexts (Part D).

AGIDE has not set out, nor will its report present, any definite answer to the often-fraught questions around ethics opened up by digital transformations. Instead it urges active, diverse and careful attention and engagement with the futures we are building today.

ON DIGITAL ETHICS

The AGIDE project did not limit itself to specific technologies, but instead intended to cover a broad spectrum of topics by allowing for their bottom-up identification and definition. We thus conceive of digital ethics at least in terms of three interrelated dimensions: (1) as a macro-ethical framework raising fundamental questions about the state of humanity, (2) as an applied ethical framework engaging with novel questions in the different areas of digital ethics, such as robotics or artificial intelligence (AI), or (3) as an assessment framework for responsible innovation of emerging or changing technologies.

Over the past three decades, as Bynum (2015) has aptly summarised, computer and information ethics has assessed the development of ICTs: first from a human-centred perspective (Parker, 1968), then from a computer-centred perspective (Moor, 1985), and more recently from an information- or content-centred perspective, focused not on hardware but on the impact of software and data across the “whole cycle of information creation, sharing, storage, protection, usage and possible destruction” (Floridi and Taddeo, 2016, p. 3; see also Weckert, 2007; Van den Hoven and Weckert, 2008; Johnson, 2009; Floridi, 2006 & 2010). “Digital ethics”, in turn, is set to engage with the “whole ecosystem created and manipulated by any digital technology [and] provides a holistic approach to the whole universe of moral issues caused by digital innovation” (Floridi, Cath and Taddeo, 2019, 11). As Kazim and Konshiyama (2021, p. 2) note, this ecosystem involves “a fusion of technologies that blur the digital, physical and biological spheres”, prompting new ethical questions around “singularity, transhumanism, and posthumanism often presented in utopian/dystopian terms”. Although these kinds of questions have always been discussed in some way (e.g. Müller, 2022), the concept of the digital ecosystem might portend a different kind of urgency accompanying present and future digital transformations (e.g. Walsh, 2022; Nagl-Docekal and Zacharasiewicz, 2022).

In addition to taking a holistic view, digital ethics can also draw on the long-established fields of general and applied ethics, as well as their traditions of thought, methods and tools. General ethics, in this regard, is “concerned with principles and practices of how humans should act in light of the problems and challenges that digitalisation poses” (Fuchs 2022, p. 5–6). It uses various different approaches, for example, virtue ethics, which sheds light on moral education (i.e. the development of virtues); deontology, including Kant’s foundational concepts such as the categorical imperative (e.g. MM, 4:421; G, 4: 402);¹ and consequentialism’s harm-benefit analyses, as well as critical theory on questions of (social) justice in (digital) society. Applied ethics, in turn, *applies* ethical approaches to more specific subfields and concrete cases (such as robot ethics, data ethics, biotechnology ethics and AI ethics), moving the ethical questions “from *what* to *how*: not just *what* ethics are needed, but *how* ethics can be effectively and successfully applied and implemented” (Floridi, 2019, p.1).

Finally, emerging technologies and innovations bring about certain transformations, with the potential to shift ways of living and thinking, as well as ethical beliefs and values (e.g. Latour, 1994; Friedman, 1997; Stiegler, 1998, Verbeek, 2008). Consequently, principles to guide digital innovation proliferate, but the relationship between digital technologies – indeed, the digital ecosystem – and societies around the world is more cyclical and complex rather than linear. As Coeckelbergh (2019, p. 139) notes, the “relationship between our devices, the world and us is neither one-dimensional nor one-way, but rather resembles a feedback-loop between technologies and everything else in the world, dynamically influencing each other”. In this context, digital ethics or principles serve as measures in how we responsibly engage with the inherent uncertainty of this feedback-loop and associated challenges (e.g. Adams and Grove, 2007).

Governments and companies, among others, have therefore striven to develop ethical guidelines, protocols and assessments to make innovation processes more responsible – encompassed in terms such as “responsible innovation” (Blok and Lemmens, 2015). A key aspect in this regard is the involvement of different stakeholders (“shared responsibility”, Von Schomberg 2013), such as ethicists, in the innovation and design process, who bear the responsibility for developing and defining “socially

¹ References to Kant’s works in the text and endnotes [I do not find any endnotes] are given parenthetically, according to the abbreviations listed here: G: Groundwork of the Metaphysics of Morals (Grundlegung zur Metaphysik der Sitten) (1785), translated by Mary J. Gregor and Allen Wood, in *Practical Philosophy*, 37–108. MM: The metaphysics of morals (Die Metaphysik der Sitten) (1797), translated by Mary J. Gregor, in *Practical Philosophy*, 353–604.

desirable products” (Von Schomberg, 2013, p. 27). Implied in such formulations is what the European Commission (EC) has defined as its *Responsible Research and Innovation* (RRI) guideline: “that societal actors (researchers, citizens, policy makers, businesses, third sector organisations, etc.), work together during the whole research and innovation process in order to align better both the process and its outcomes with the values, needs and expectations of society” (Delaney *et al.*, 2020, p. 8). This emphasis on “working together” in the entire life cycle of technology innovation puts newfound emphasis on the (new) methods to achieve the aims of the RRI, including methods that aim to analyse practical ethical issues that arise in the planning stages of new technologies, during technology development and in assessing their impacts on society (e.g. Reijers *et al.* 2018). This also includes using theoretical future scenarios and the envisioning of potential futures as ethical assessment strategies, such as the techno-ethical scenario approach, the ETICA approach² or enhanced ATE (anticipatory technology ethics) approach (e.g. Umbrello *et al.*, 2023).

ON THIS REPORT

Guided by a broad and open understanding of digital ethics and associated methods, AGIDE brought together international scholars and experts to engage with their different, situated perspectives on digital ethics. The resulting variety shows that while ethical considerations may appear

similar “from a distance”, or in the abstract, in practice, digital ethics often reveal a complexity of different aspects, interpretations and narratives “on the ground”.

The AGIDE report is organised around four parts from A to D: Part A provides the theoretical groundwork, Parts B and C set out our empirical and analytical efforts and Part D summarises the implications of the project as a whole. **Part A** introduces the theoretical framework of the project and the tension between universalist and localised values inherent in a project on digital ethics that strives to move away from Eurocentric stereotyping and dominant views from the perspective of what is called “the West” or the “Global North”. In **Part B**, we set out a largely thematic analysis of the expert interviews and workshops that comprise the empirical basis and the general findings of the project. The reader will find in them a diverse range of ideas and notions of digital ethics, topics of interest and themes of concern for the present and future of digital ethics (in practice). **Part C** offers one way of synthesising these findings into a matrix and demonstrates its flexible usability as a scheme for context-specific and situated narratives about the uses and development of digital technologies. As such, it offers one possible way of systematising the variable data generated through this project, while also opening up ample avenues for future research. Examples of such avenues and general implications of the project as a whole are summarised in our conclusions in **Part D**.

² Ethical Issues of Emerging ICT Applications, following a 2009-11 European Commission-funded project (Grant agreement ID: 230318, see also <https://cordis.europa.eu/project/id/230318>).

PART A: DIGITAL ETHICS – WITH A FOCUS ON DIFFERENCES

As a theoretical introduction to the issues to be discussed in the later parts of this report, the following Part A explores the complex relations, and at times tensions, between global digital ethics approaches and localised, situated perspectives to provide a framework for the empirical analyses of AGIDE.

UNIVERSALIST APPROACHES TO DIGITAL ETHICS

Recent years have seen a surge in academic literature on digital ethics, many proposing different sets of “principles” (guidelines, recommendations, frameworks, etc.) for digital ethics. Such principles have been developed within public and private sectors at national, regional, international and global levels. Most of them, be it explicitly or implicitly, see digital ethics everywhere in the world as facing similar challenges, potentially offering similar answers and following similar notions of the “good”. Abstract notions such as “fairness”, “accountability” and “transparency” – widely known as the “FAT principles” – have become common denominators in most ethics documents around the globe, as will be further discussed below. A characteristic shared by these documents is that the number of ethics principles they propose has increased over time.

Fjeld *et al.* (2020, p. 66), for example, mapped the landscape of artificial intelligence (AI) principles on a global scale, identifying common themes. Based on this analysis, the authors concluded that existing ethics and rights-based principles have started to converge in recent years. However, they further argue that there is “a wide and thorny gap between the articulation of these high-level concepts and their actual achievement in the real world”. To conclude, they suggest that more work needs to be done on identifying variations within the themes, especially with regard to particular geographies or stakeholder groups. Moreover, the modes of production of these documents need to be considered. Roche, Wall and Lewis (2022) analysed the distribution of AI ethics and policy documents into world regions/countries, finding that voices from the Global North, from the United States (US) and Europe more specifically, produced a dominant majority of the existing documents. In addition, the analysis of prominent

keywords in these documents revealed that there is little reference to underrepresented populations and “low resource”, “low/middle income” and “vulnerable” groups. All this research points to the necessity for ongoing work on digital ethics to look *beyond* dominating voices, proliferating themes and principles, and *into* underrepresented voices and overlooked themes.

Other authors question ethics principles on a more fundamental level. Munn (2023, p. 870), for example, describes the gap between high-minded principles and technological practice as dangerous since the translation from complex social contexts to technical rulesets is non-trivial: “In a zero-sum world the obsession with AI principles is not just useless but dangerous in funnelling human and financial resources away from more productive approaches”. Van Maanen (2022) problematises the term “ethics” itself by arguing that the strong focus on ethics obstructs our view of alternative tools and methodologies to engage with the tech sector, such as human rights or data justice approaches. He therefore calls for (re)politicising ethics itself. Emphasising the importance of empirically informed analyses of techno-political practices, he argues for an understanding of ethics as a collective and interactive endeavour, acknowledging “the conflictual character of doing political work”. Labelling such practices “politics” (e.g. “data politics”), rather than “ethics”, would make them less susceptible to being washed away in the public relations machinery of the tech sector – widely framed as “ethics washing”.

It is against this background of critical reflections on the proliferation of digital ethics discourse that the AGIDE project set out to conduct its research. In the next step, we map the wide landscape of global ethics principles. Below that, we dive deeper into the complex relations between shared ethics principles and “real-world” or more “applied” practices.

EXISTING SETS OF PRINCIPLES

Existing sets of principles look strikingly similar, mostly listing between three and ten general principles at a high level of abstraction. The United Kingdom (UK), for example, published the *UK Data Ethics Framework* in 2018, which

is intended to guide responsible data use in public sectors. The three broad ethics principles mentioned in the framework are fairness, accountability and transparency, closely resembling the “FAT paradigm”. In the framework, “fairness” aims to eliminate a project’s potential for unintended discriminatory effects for individuals or social groups by mitigating biases and ensuring that the project’s outcomes “respect the dignity of individuals, are just, non-discriminatory, and consistent with the public interest, including human rights and democratic values” (Government of the UK, 2020). “Accountability” means that there are effective governance and oversight mechanisms in place for every project; “public accountability”, in turn, that the public, or its representatives, can exercise oversight and control over actions taken by governmental and other public institutions. “Transparency” refers to the way information about projects, processes and actions is made publicly available for inspection. Similarly, the Organisation for Economic Co-operation and Development (OECD) published a report in 2020 formulating *10 Good Practice Principles for Data Ethics in the Public Sector* (OECD, 2020). In addition to openness and accountability, which are also mentioned in the *UK Ethics Framework*, the OECD’s principles briefly mention “inclusiveness”, as well as individuals’ and collectives’ “control over their data”. Accordingly, despite different formats and wordings, both documents focus on similar principles for responsible data use by governments and public sectors, revolving around openness, control and accountability.

Since 2018, the number of published ethics principles, especially in the form of guidelines, has grown significantly, not least regarding AI. Fjeld *et al.* (2020) from the Berkman Klein Center reviewed 36 documents on AI principles from all over the world, namely Latin America, East and South Asia, the Middle East, North America and Europe. They stem from governments and intergovernmental organisations (e.g. *AI in the UK*, *AI for Europe*, *AI in Mexico*, the *German AI Strategy* (e.g. see: Wendehorst and Woopen *et al.*, 2019) or *AI Principles* by the Japanese and the Chinese government, *OECD Principles on AI*, or *G20 AI Principles*), companies (e.g. *Microsoft AI Principles*, *AI at Google: Our Principles*), professional associations (*Montreal Declaration* by the University of Montreal or the *AI Code of Conduct* by the Chinese AI Industry Alliance), advocacy groups and multi-stakeholder initiatives (*Toronto Declaration* by Amnesty International, *Human Rights in the Age of AI* by Access Now or *Universal Guidelines for AI* by the Public Voice Coalition). Across all of these documents, Fjeld *et al.* identified a total of eight key themes: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control over technology, professional responsibility and promotion of human values.

The eight key themes comprise the following aspects in detail (our summary):

1. **Privacy** relates to control over user data and consent, privacy by design, restricting processing and the right to rectification and erasure.
2. **Accountability** includes impact assessment, auditing requirements, the ability to appeal, environmental responsibility, remedy for automated decision-making, access to a monitoring body and the possibility of redress.
3. **Safety and security** refer to safety and reliability, predictability and security by design.
4. **Transparency and explainability** encompass open-source data and algorithms, notifications when AI is at play in interactions or decision-making processes, reporting requirements, the right to information and open procurement for governments.
5. **Fairness and non-discrimination** include the prevention of bias, inclusiveness in design and impact, representative and high-quality data (including accuracy of the data), and equality.
6. **Human control of technology** refers to review of automated decision-making and the ability to opt out of automated decision-making processes.
7. **Professional responsibility** links to multi-stakeholder collaboration, responsible design, consideration of long-term effects, accuracy and scientific integrity.
8. **Promotion of human values** refers to human values and human flourishing, access to technology and technology “leveraged to benefit society”.

These eight key themes provide a useful high-level snapshot of worldwide AI principles. The authors of the review observed that while there was considerable variation in the key themes in older documents, more recent publications tend to cover all eight themes, suggesting that AI principles are beginning to converge. A possible explanation for this, as they argue, is that most of the documents are drafted by a relatively small group of people, who know each other and are moving towards a consensus. This is confirmed by earlier research by Jobin, Ienca and Vayena (2019), who also mapped the global landscape of AI ethics guidelines. Having analysed 84 documents, mainly from the Western hemisphere (see full list of documents in Jobin, Ienca and Vayena, 2019, pp. 4–5), they concluded that convergence emerged around five central ethical principles: transparency, justice and fairness, non-maleficence, responsibility and privacy. However, Jobin, Ienca and Vayena (2019, p. 1) also observed a “substantive divergence in relation to how these principles are interpreted; why they are deemed important; what issue, domain or actors they pertain to; and how they should be implemented”. Privacy, for example, is interpreted differently in countries like China

compared to its meaning in the European Union (EU) (see Arora, 2019a). Moreover, perceptions of safety and security vary between regions, as do understandings of discrimination. What constitutes discrimination is often legally determined by states and can differ based on cultural context – the concept of caste in India or gender in the Middle East and North Africa can be seen as prominent examples of this (Arora, 2019a; 2019b). This showcases the need for further research on the contextuality, situatedness and cultural variety on the ground, as we argue below.

Adding to the documents reviewed by the Berkman and Klein Center, we analysed two more recent and important documents on AI ethics. The first is corporate and multinational: the PwC (Pricewaterhouse Coopers) *Ethical AI Principles* by the World Economic Forum published in 2021 (Golbin and Axente, 2021). This document lists nine AI ethics principles and distinguishes between epistemic and general principles. The two sets of so-called “epistemic principles” are interpretability (explainability, transparency, provability) and reliability, robustness and security. They are considered prerequisites for AI to be ethical in the sense of representing the “conditions of knowledge that enable organizations to determine whether an AI system is consistent with an ethical principle” (Golbin and Axente, 2021). The seven general principles include accountability, data privacy, lawfulness and compliance, beneficial AI, human agency, safety, and fairness. Out of these principles, the principle of “beneficial AI” – ensuring that AI promotes the common good such as sustainability, cooperation and openness – is worth highlighting here.

The second document we analysed further is the *United Nations Education, Scientific and Cultural Organisation’s (UNESCO) Recommendation on the Ethics of AI*, adopted in 2021 and published in 2022. Compared to the other documents, the *UNESCO Recommendation* puts a stronger focus on human dignity, diversity and inclusiveness. It comprises ten core principles altogether: proportionality and *do no harm*, safety and security, the right to privacy and data protection, multi-stakeholder and adaptive governance and collaboration, responsibility and accountability, transparency and explainability, human oversight and determination, sustainability, awareness and literacy, and fairness and non-discrimination. The first principle, “proportionality and *do no harm*”, deserves particular attention, as it suggests that the use of AI systems must not go beyond what is necessary to achieve a legitimate aim. Accordingly, it stages itself as “the first international normative instrument that contains a provision against using AI systems for social scoring and mass surveillance purposes” (UNESCO, 2022, p. 8). The notion of “mass surveillance” may also include surveillance for “the good” in the sense that welfare states may sometimes choose continuous surveillance as a measure to ensure that citizens are protected, safe and have access to welfare state institutions, especially in the Global South.

Additionally, principle eight in the *UNESCO Recommendation* refers to sustainability, and the United Nations (UN) Sustainable Development Goals more specifically, to which AI technologies should adhere. Notably, neither of the reports on AI ethics discussed above by Fjeld *et al.* (2020) and Jobin, Ienca and Vayena (2019) referenced “sustainability” as one of their dimensions, which underlines the growing alertness to the importance of the matter on a global scale. Finally, principle nine in the *UNESCO Recommendation* suggests the promotion of public understanding of AI through open and accessible education, civic engagement, digital skills and AI ethics training, media coverage and information literacy.

In addition to its ethics principles, the UNESCO Recommendation (2022, p. 16–17) formulates “actionable policies” to move beyond high-level principles and work towards practical strategies. The policy areas mentioned include ethical governance and stewardship, economy and labour, data policy, health and social well-being, education and research, and gender, as well as environment and ecosystems. Furthermore, UNESCO member states are provided with “actionable resources” to implement the recommendation, such as the “readiness assessment methodology” to identify the status of preparedness or the “ethical impact assessment” to identify the potential impacts of an AI system. This, in turn, might be interpreted as a first important step towards acknowledging national contexts, cultural differences and divergent political and economic systems engaged in practical attempts to put a global recommendation into practice. It is to this matter – the differences beneath general principles – that we turn next.

DIFFERENCES BENEATH THE SURFACE

Intercultural digital ethics (IDE) is a subfield of information and digital ethics research which examines ethical issues related to digital technologies from different cultural and social perspectives (cf. Hongladarom, 1999; Capurro, 2005; Ess, 2006; Hongladarom and Ess, 2007; Capurro, 2008). This field of research addresses how digital technologies affect different groups and the challenge of developing a global and “cosmopolitan” (e.g. Ess 2020; Jasanoff, 2013) yet pluralistic intercultural digital ethics approach (Ess, 2006) that reflects heterogeneous cultural values at the same time as cross-cultural governance cooperation (e.g. Floridi, 2019; Taylor, Floridi and van der Sloot, 2017; Couldry and Mejias, 2019). The most important issues in IDE include the study of ethical systems based on non-“Western” traditions such as Buddhism, Confucianism, Ubuntu and others (e.g. Wong, 2012; Reviglio, Alunge 2020; Berberich, Nishida, & Suzuki, 2020; Chaudhary, 2020).

Having analysed shared sets of global ethics principles, as discussed above, AGIDE chooses to focus on *differences*: both differences *between* global regions and also differences *within*

countries, considering that not only cultural values but also socio-economic and other salient differences are at play here. Fjeld *et al.* (2020, p. 5) argue that, more generally, “principles should be understood in their cultural, linguistic, geographic, and organizational context”, whereby “some themes will be more relevant to a particular context and audience than others”. In this interpretation, ethics principles should be seen as the smallest common denominator on which everyone, or a majority, can agree. Global ethics principles and recommendations, such as the *UNESCO Recommendation*, therefore need to be implemented and translated into specific contexts, taking into account local circumstances and particularities. Consequently, for instance, experts are currently appointed to national advisory boards to help with the implementation of the *UNESCO Recommendation* into national contexts (e.g. Austrian Advisory Board, 2024).

LOCAL CONTEXTS AND APPLICATIONS OF DIGITAL ETHICS PRINCIPLES

Differences become visible when taking a closer look at how different communities understand the various principles. For example, the *FAIR & CARE Principles* by the Global Indigenous Data Alliance look (Carroll *et al.*, 2022), at first glance, very similar to the sets of principles discussed above. After restating the well-known standard of FAIR (findable, accessible, interoperable, reusable), they feature four principles: “collective benefit”, “authority to control”, “responsibility” and “ethics”. While this does not seem to differ fundamentally from other general digital ethics principles, a closer look at the explanations provided reveals different interpretations and different levels of priority regarding Indigenous Peoples’ specific situation and needs (Carroll *et al.*, 2022):

1. **Collective Benefit:** Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.
2. **Authority to Control:** Indigenous Peoples’ rights and interests in Indigenous data must be recognised and their authority to control such data be empowered.
3. **Responsibility:** Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples’ self-determination and collective benefit.
4. **Ethics:** Indigenous Peoples’ rights and well-being should be the primary concern at all stages of the data life cycle and across the data ecosystem.

This illustrates that, below the surface, concepts such as “responsibility” or “fairness” are open to very different interpretations and enactment – such as stressing autonomy, equality or harmony. “Privacy” can also refer to an array of different things, depending on location in the world

and the social group(s) in question. Cho, Rivera-Sánchez and Lim (2009), for example, investigate internet users’ perceptions and behavioural responses concerning online privacy using a representative sample of multi-national internet users from Bangalore, Seoul, Singapore, Sydney and New York. Their results show that individual differences (age, gender, education and personal income), but also nationality and national culture significantly influenced internet users’ privacy concerns. For example, internet users from “high individualism” countries were more likely to be concerned about potential privacy intrusions than those from countries where individualism plays a lesser role. This might relate to the fact that a “collectivist” tradition respects the privacy of an individual from the perspective of the community (see also Arora, 2019b); that is, asking rather how would respecting the privacy of an individual empower the individual to contribute to the good of the community. Moreover, intricate relations between duties and rights need to be considered in related cultural traditions, the details of which go beyond the scope of this report. However, Cho, Rivera-Sánchez and Lim (2009) conclude that a more comprehensive and holistic analysis of online privacy is needed, given that the way in which people perceive and respond to online privacy is also affected by macro-level, national and cultural differences alongside political ones.

Socio-political and cultural differences also impact the way data protection is negotiated and regulated in different countries, regions and supranational organisations. In the EU, for example, heated negotiations over the General Data Protection Regulation (GDPR) revealed marked differences between the member states and their positions towards privacy and data protection, which were deeply rooted in different histories, political regimes, social values and economic cultures (Mager, 2017). Moreover, normative positions in respect of (new) digital technologies vary significantly between different societal groups, as has been shown, for example, in the context of COVID-19 contact tracing apps (Lucivero *et al.*, 2022; see also Lanzing, 2021). It is, therefore, widely recognised at this point that formulating principles for digital ethics requires contextualisation, specifically including a cultural dimension. The relative weight, for instance, given to one aspect over another will likely vary considerably. This focus on differences needs to be taken into account when implementing shared ethics principles, but to no lesser degree in their initial development. While many of the global ethics principles were developed in countries of the so-called West or Global North, as discussed above, there are exceptions to that as well: the *UNESCO Recommendation for AI Ethics* and the *World Health Organisation’s (WHO) Guidance on Ethics and Governance of AI for Health*, for example, tried to include experts from different parts of the world in their development processes according to stakeholders involved (in the

WHO guidelines, the experts are listed, see WHO, 2021). Nonetheless, it stands to reason that these shared principles also need to be broken down and made more concrete in order to become operational in specific geographical and cultural contexts. As noted above, in the case of the *UNESCO Recommendation*, this is attempted with the help of actionable policies and national advisory boards.

To be able to contextualise shared principles and make them actionable in practice, the concept of “situatedness” is a helpful analytical tool. We turn to the meaning and utility of considering the role of “situatedness” in the following section.

WHY SITUATEDNESS MATTERS

The term “situatedness” has become a common notion in Science and Technology Studies (STS) to place the analytical focus on differences in social, cultural, political, economic and institutional positionality (Pinch and Collins, 1982; Haraway, 1988; see also Butler, 1990; Thompson, 2002; Pfotenhauer *et al.*, 2022). As a response to generalising and globalising tendencies, the term situatedness allows an emphasis on political contestations, historical contingency and interpretive flexibility (Pinch and Bijker, 1984). Haraway (1988, p. 589) argues:

... for politics and epistemologies of location, positioning, and situating, where partiality and not universality is the condition of being heard to make rational knowledge claims. These are claims on people's lives. I am arguing for the view from a body, always a complex, contradictory, structuring, and structured body, versus the view from above, from nowhere, from simplicity.

This “view from a body” includes, in Haraway’s view, building an “earth-wide network of connections” and developing the ability to translate different, localised forms of knowledge among different and unequally powerful communities (Haraway 1988, p. 580). Reflecting on situatedness as an analytic perspective with a long history, Pfotenhauer *et al.* (2022, p. 8) argue that it “has opened up spaces for a normative critique of hegemonic power structures and colonial tendencies that threaten to erase epistemic and political diversity”, referring to work by Visvanathan (1997), Harding (2006), TallBear (2013) and Prasad (2014), as well as Tsing’s (2005, p. 38) idea of constituting “meaningful diversity”. This relates to newer debates on “epistemic justice” and the idea that some forms of knowledge are more valued than others due to the positionality of the knower. In her book on the topic in the context of knowledge production, Fricker argues that we cannot consider any epistemological issue without also considering the “ethical and political aspects of our epistemic conduct” (2007, p. 2). Zhang and Datta Burton (2022), in turn, engage with aspects of power

in the production and governance of scientific knowledge by calling for its decolonialisation.

To give some examples, we turn to STS discussions on digital ethics and how hegemonic discourses obstruct our view of situated approaches and epistemic justice. Pfotenhauer *et al.* (2022) argue that companies like Google or Facebook have enthusiastically ab/used ethics approaches to deal with the “corrosive effects of their scale” as a form of self-critique that seeks refuge in universal principles and post-hoc institutional fixes, obfuscating more fundamental questions about how digital platforms should be developed or governed. This corresponds to critique mentioned earlier by van Maanen (2022), which calls for a „(re)politicisation of ethics itself” to open up the view on fundamental rights and justice approaches so as to capture the technopolitical practices at work. Ricaurte (2022), for example, criticises an AI ethics “for the majority of the world”. She argues that hegemonic AI exerts global violence through three epistemic practices: (1) datafication in terms of extraction and dispossession, (2) algorithmisation in the sense of mediation and governmentality and (3) automation linked to violence, inequality and displacement of responsibility. Ricaurte (2022, p. 726) concludes that these “articulated epistemic mechanisms lead to global classification orders and epistemic, economic, social, cultural and environmental inequality”.

Moreover, Pfotenhauer *et al.* (2022) argue, corporate hegemonic discourses, such as the “scalability zeitgeist” (the “fixation on ‘scaling up’”) spill over to other sectors and become “key ordering logics” of current innovation and public policy at large – flattening out technological diversity and socio-cultural contexts even more. Accordingly, there is a growing body of STS research focusing on national and cultural differences in internet policy and governance, mostly from the “Western” academic world (Mager, 2017 & 2018; Aho and Duffield, 2020; Barais and Katzenbach, 2022; Guay and Birch, 2022). What is still lacking is a (more) comprehensive understanding of how “situatedness” – in terms of cultural values, socio-economic positionings, political regimes, centre/periphery and so on – plays into digital ethics. Voices from the Global South and other marginalised geographies and communities in particular need to be taken into account when trying to understand and realise contextuality in global ethics discourses.

The AGIDE project therefore tries to place its focus on *differences* by including research participants from all over the world in an attempt to make global ethics less homogenic and more sensitive to, and situated in, the real world *on* the ground and *from* the ground *up*. Put again in Haraway’s (1988, p. 590) terms: “The only way to find a larger vision is to be somewhere in particular.”

The following Part B comprises an overview of the diversity of topics discussed with project participants through the empirical engagements of the AGIDE project. Throughout,

the reader will find outlined some of the issues in digital ethics today as they relate to specific regional, social, political or economic contexts, as well as key overarching issues. Therein, we also find ample starting points for further engagements and considerations. Part C, thereafter, provides a means of systematically synthesising the breadth of

topics, themes and futures discussed in a shareable matrix that corresponds to our theoretical underpinnings discussed above. It also provides the basis for meaningful and situated ethical narratives, which we found to be one apt means of integrating our findings without simply aligning or “typecasting” them wholesale.

PART B: THE AGIDE PROJECT – EMPIRICAL FINDINGS

THE EMPIRICAL ANALYSES – INTRODUCTION

Digital transformations are changing the way we live and the way our societies and our economies function, as well as global power relations. They have brought about an unprecedented degree of global interconnectedness, enhancing efforts to formulate transnational ethical guidelines for dealing with emerging digital technologies. The AGIDE project, started in 2022 and spearheaded by the Austrian Academy of Sciences, primarily sought to explore the cultural dimensions of digital ethics and global innovation – that is, how technologies are transforming our societies and global structures – by taking off its “Eurocentric glasses”, embracing differences, and going beyond familiar stereotypes or clichés. To this end, the Austrian Academy of Sciences cooperated with other academies of sciences from all over the world. AGIDE aimed to embrace the socio-cultural variety of perspectives and to further explore differences and similarities without forcing uniformity or consensus. The overarching objective was to explore cultural and other dimensions of note for digital ethics and delving beyond discussions already existing in the literature. Whereas AGIDE is foremost a research project and not geared towards specific applications, its insights may be deemed fruitful in a variety of contexts – from embedding ethical values in technology throughout the product life cycle to facilitating international efforts at legal harmonisation.

THE AGIDE PROJECT – GENERAL METHODOLOGY

The AGIDE project employed a qualitative methodological approach, encompassing semi-structured expert interviews, a workshop series with invited talks and a “use case” discussion to gather empirical evidence and understand diverse global perspectives on digital ethics (see Figure 1).

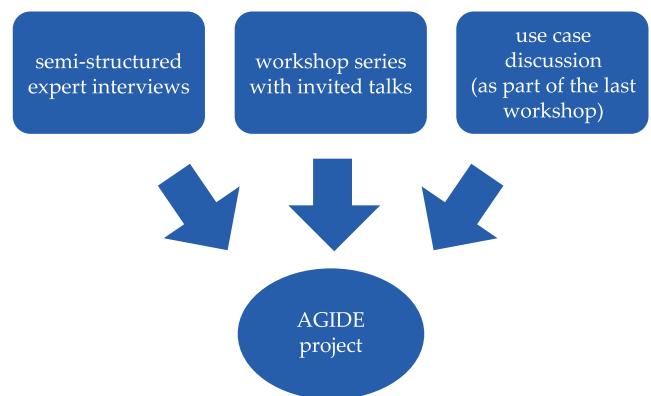


Figure 1: Methodology of the AGIDE project [source: Working Group Members]

Following a constructivist perspective, our qualitative approach views people, social systems and technological development as interrelated and emphasises the relevance of stakeholder involvement and the value of concrete, situated knowledge. Identifying the boundaries and situated limitations that come with positivist positions on a single, universally shared and neutral reality has a long history in the social sciences and humanities (Haraway, 1988; Firestone, 1993), especially when studying people and their lived experiences. In turn, researchers engage in data collection as active participants and consider knowledge creation as multi-faceted, mind-dependent and necessarily contextualised and partial. In other words, such an approach appreciates the situated positions from which knowledge can be produced (Hartsock, 2019) and its interpretive diversity with recourse to their inherent theories about the world (Mayring, 2007).

Given that the AGIDE project sought to embrace this diversity of perspectives and explore differences, such an approach is uniquely suitable. Our data analysis thus aims to provide a global picture of diverse attitudes towards technology, as well as “narrative snapshots” (see Part D) of particular situations and approaches to their evaluation. AGIDE purposely refrains from presenting our findings in a purely abstract manner as this would put the focus on commonalities rather than differences. However, it is important to note here that it is not the project’s aim to

portray individuals as *representative* of particular countries or cultural regions, nor to *speak for* certain cultures or regions. As noted in Part A of this report, the results presented offer insight into *situated perspectives* which arise within a particular context, time and space, where they are situationally constructed and negotiated.

Finally, in Part C, we map findings from the interviews and workshops onto an “matrix of digital ethics narratives” as a means to distil, analyse and invite further exploration and discussion of the AGIDE results. The mapping was derived from inductive coding based on the empirical findings and discusses digital ethics with regard to eight interrelated dimensions that form a matrix to abstract from and summarise our findings. From the matrix we derive potential *narratives of digital ethics*. Narratives, in that sense, are stories that are told repeatedly, consisting of a series of events that are selected and arranged in a particular order, often including central characters (protagonists, antagonists), a conflict and a plot. In the presented narratives, we use the dimensions and their associated approaches, as identified in the matrix, as key narrative elements, story beats and actors. In doing so, we exemplify not only differences within a certain region (case study: the EU) and population (case study: India) but also how situated differences can be made to matter on a more abstract level.

Based on all findings discussed in Part B, we present overarching “lessons learned” and possible ways forward, for further discussion and future research in a Part D on the implications of the AGIDE project writ large.

THE AGIDE PROJECT – LIMITATIONS

The research goal of the AGIDE project was not to come up with generalisable findings or scientific “truths”. Rather, the primary goal of the empirical research was to collect both shared understandings of and differences in the notion of digital ethics, situated within their cultural contexts. A relatively low-threshold empirical methodology enabled a broad range of experts to take part in the project, supported by partnerships with global science academies. Therefore, interviews were conducted remotely and workshops were offered in a hybrid (in-person and digital) format to invite diverse participation, independent of any limitations on travel.

However, as with every research project, some limitations must be discussed and reflected on critically:

- **Contextual Limitations:** Findings are always influenced by the specific context in which a project is conducted, including the selection of participants and the nature of the questions asked. For AGIDE, this involves the fact that interviews had to be conducted in English and that interviewees were identified through multiple rounds of sampling (see below) that could include a bias in favour

of a small “bubble” of international experts. We might also assume a tendency for views expressed to reflect stereotypical or “hegemonic” positions (Collins, 2008) that have become dominant in literature, expert circles, publics and so on. The project therefore does not exhaust the full range of approaches to digital ethics which might be found across the world. Experts from around the world were asked about dominant opinions regarding digital ethics in relation to the regions in which they currently live (or previously lived) or have specific expertise on. Yet the interviewees, while providing valuable insights, do not represent or *claim* to represent these regions. They do, however, trace current and historical discourses and trends to which they have become witness. This, in turn and across the entire interview data gathered for this project, allows for insightful conclusions which are further aggregated in Part C of this report.

- **The Expert Voice Fallacy:** The “expert voice fallacy” points to the idea that scholars, activists, and all those generally considered “experts” in their fields might earn or develop reputations by taking very original or indeed extreme positions, widening the gap between more localised lived experiences and the so-called experts’ positions. While this may always be the case, we have tried to account for this fallacy by way of triangulation, i.e. including a large number of interviewees and conducting multiple rounds of workshops.
- **Generational Gap:** Having interviewed generally established experts, a notable potential gap might exist between the generations of professionals in the field of digital ethics and people who are younger. The perspectives of the latter are largely absent.
- **Open-Endedness:** Digital ethics may, as noted above, refer to a great number of past or current digital technologies. To foster a broad debate, yet tailor it to the specific understandings and frameworks of individual interviewees, AGIDE chose not to define which technologies to consider in the debate surrounding digital ethics. We did not provide a shared mental model or guided list as part of our open-ended research design. Interviewees were asked which digital technologies currently play the largest and most impactful role in society and were asked to specify their reasons why. This open-endedness was considered an asset in answering the overall research question and in line with our situatedness approach, but future research might opt to focus in more granular detail on certain technologies, platforms or issues.
- **The Evolving Nature of Digital Ethics:** Digital ethics is a rapidly evolving field, and the ethical challenges posed

by digital technologies continue to emerge and evolve. The findings obtained from the AGIDE project may be considered preliminary and capture only a momentary snapshot of current understandings and perspectives on digital ethics.

Overall, it is important to consider the findings of the AGIDE project in light of its goals and of these limitations. AGIDE provides a broad starting point for further research, exploration and refinement. Both its findings and its ambition highlight the constant need for continued investigation, engagement with diverse stakeholders and ongoing dialogue to gain nuanced understandings of the diverse approaches to digital ethics worldwide.

In the following, we present findings first from our expert interview analysis with 75 participants, then from two input workshops that were held as part of the project, and finally of a third workshop on future(s) scenarios.

THE AGIDE INTERVIEWS: ANALYSIS

THE AGIDE INTERVIEWS – INTRODUCTION

Our qualitative approach allowed us to tap into the situated knowledge of so-called “experts”: individuals with privileged access to information on and about digital ethics (Meuser and Nagel, 2009). While expert interviews are fundamentally considered an effective method for gathering more specific information about a particular subject matter or domain, careful reflection is still required. Bogner, Littig and Menz (2009, p. 1) note that “issues on what constitutes an expert, the differences between the various forms of expert interviews and their role in research design” quickly arise. The AGIDE project did not define “experts” purely by practice of a specific profession with which such expertise might be institutionally associated (such as ethicists in the academy). Therefore, we have opted to define “experts” as people who have privileged knowledge on and about digital ethics, and who play an active role in structuring and defining the relevant issues and problems related to it. The pool of experts for the interviews was based on a combination of (initial) purposive sampling and (subsequent) snowball sampling, drawing on the expertise of the steering committee and academies of sciences around the world. Purposive sampling of potential experts followed considerations of the general domain of individuals’ proven expertise in digital ethics alongside considerations of diversity. Additional selective sampling was employed to cover remaining gaps in knowledge and expertise and based on collective decision-making within the project’s steering committee.

In total, the AGIDE project gathered data by conducting 75 qualitative interviews with globally renowned experts between October 2022 and November 2023. The interviews were semi-structured using a guiding questionnaire

(Silverman, 2021), developed by the AGIDE working group according to the project’s goals, digital ethics dimensions and responsible innovation methods (see Part A). Experts were chosen from across different world regions, disciplinary backgrounds, and expertise in a variety of fields inside and outside of academia, spanning to include activism and industry. While achieving representativeness was not a goal in this qualitative research design, the working group aimed to cover a broad range of experts in order to answer the main research question as thoroughly as possible. As such, the experts spoke from situated and necessarily privileged “standpoints” embedded in diverse cultural contexts (Collins, 2008).

In this regard, the term “culture” therefore not only extends to include factors such as geographical location, language or religion, but also disciplinary cultures that may greatly influence how digital ethics are discussed and the norms and values that are considered most relevant. Experts had disciplinary backgrounds from anthropology, comparative literature, computer sciences, economics, ethics, law, medicine, media studies, philosophy, sociology, psychology and others. Gender parity was reached amongst interviewees. The interviews were analysed using thematic analysis: first, initial codes were generated across the dataset of interviews which were then subsumed as expressing certain themes (Braun and Clarke, 2006). The themes are represented by headlines in the overview below, with more detailed insights gathered from the data included in the summaries.

MEANINGFUL DIFFERENCES IN DIGITAL ETHICS

Our interview data supports the notion that “culture” is a multifaceted term and concept, not least in the realm of digital ethics. To fully explore cultural similarities and differences in this field, this complexity must be analytically accounted for. As interviewees pointed out repeatedly, there is no one “culture” in the countries they speak on. Rather, different citizen groups or groups of people may experience digital ethics or “the good” vastly differently, depending on “intersectional” factors (Crenshaw, 1989) present within the country or region they inhabit. These intersectional factors may include, but are not limited to, age, gender, ethnicity, educational background, income levels and origins – often tied to access to digital infrastructures or lack thereof. These factors may greatly impact how people experience and evaluate the digital world around them. In turn, they influence which aspects of digital ethics seem most relevant to them. Moreover, there are many and intersecting “collectives”³ – in terms of “global(ised) cultures” – connecting people across regions and in different

³ This terminology loosely refers to the work of Ludwik Fleck (1981[1935]).

places on questions of digital ethics. For example, software engineers (or social scientists or philosophers) across the world may have views more similar to each other rather than to other groups and citizens living in the same country as them. As cultures intersect, it is never possible to be a member of only one “culture” or one collective – identities too are multifaceted and present only and necessarily “partial connections” (Strathern, 2004). Accordingly, people’s views on digital ethics are influenced by such partial connections and may even take on differing views on digital ethics within themselves, depending on the position from which they are speaking at a given moment.

Another key overarching finding that emerged from the AGIDE interviews is that many key concepts in digital ethics – such as freedom, autonomy, safety, etc. – are used by people with different (cultural) backgrounds and are framed as being of particular importance. However, when analysing in detail what these concepts mean in different contexts, a range of different, historically and culturally anchored meanings emerge behind these words. Let us take “freedom” as an example: From a Eurocentric or Western-centric perspective, freedom in the context of digital ethics may include *freedom from* prosecution and surveillance, *freedom of* choice in one’s actions, and/or the *freedom to* move safely in digital spaces while one’s data is protected (which ties into other values, such as “data privacy”). It might also mean the freedom to maximise your opportunities when finding employment, which may be negatively impacted by algorithmic profiling and thus impinge on *freedoms* (plural). To give another example, freedom – viewed from an indigenous data sovereignty perspective – is about having the jurisdiction and control over data so that the indigenous community can define, collect, own and apply their data in a uniquely indigenous way, without colonial interference.

These definitions or aspects of “freedom” already cover a broad range of factors. However, when looking into definitions or understandings of freedom that are not Western-centric, all of the dimensions above may still be considered relevant while novel ones might gain precedence. Following familiar stereotypes or clichés that were, at least partially, supported by AGIDE interview data and reproduced by the experts interviewed, a stronger notion of community-orientation was expressed from non-Western standpoints. In turn, individuality was associated more strongly with Western ideas or ethical principles. In more community-oriented contexts, freedom can never (only) entail *freedom for* the individual because we are all fundamentally tied to each other. Freedom can therefore never merely be something a single person chooses for themselves or is granted on an individual level. With regard to data privacy, one individual choosing to share data – something a Western-centric lens might label as the “freedom of individual choice” – takes on an entirely different

meaning if one’s individual data also allow conclusions about one’s family, friends and other community members to be drawn. If the value of such ties is much higher than one’s individual needs and rights, the freedom of individual choice needs to be understood and interpreted in the context of other values. Freedom, accordingly, could entail something altogether different, such as more opportunities for the community at large, for example, by providing better digital infrastructure(s) or governance measures that allow communities to self-govern their data (which in turn may potentially enhance the choices open to individuals). Consequently, a key finding from the AGIDE interviews is that variation in the meaning, definitions and understandings of words and concepts central to digital ethics is significant. **A culturally sensitive approach to analysing digital ethics can, therefore, never merely compare different guiding principles on the surface (content) but must go on to explore the more profound meaning (hermeneutics) behind these concepts, grounded in their material contexts (situatedness).** One way to operationalise such an approach can be found in the “matrix of digital ethics narratives” which illustrates the narratives expounded in Part C of this report.

That being said, there were values in digital ethics assigned by interviewees to specific cultural groups or backgrounds. Some experts stated that Shintoism, as well as Māori culture, and many other indigenous cultures, see all things as being infused with spirit or life force, and this may include material digital technologies, such as robots. This could lead to a more positive stance towards care robots and/or a different interaction with technologies from a Shinto perspective. Buddhist ethics, as discussed by some experts, may put emphasis on digital ethics to avoid harm and suffering, which could entail opting to regulate the digital space in ways that avoid possible outcomes that could cause current and future harm. Digital ethics in Confucianism, according to some of our experts, could entail regulating digitalisation in a way that utilises it to improve society and foster a more harmonious global coexistence. In this context, it should be noted that interviewees stressed that ideals such as these are not only relevant for those closely following these religious or philosophical practices. Instead, they emphasised the significance of such ideas and ideals as principles that may (more or less loosely) connect different regions, communities or groups.

Cultural differences can also call for very specific understandings of digital governance: For example, some indigenous communities have voiced their wish to honour a principle of not recording and disseminating local rituals and practices, a wish made increasingly difficult through the wide availability of smartphones and social media.

Taken together, reference to culture or to certain collectively shared opinions, meanings or ideas about the world *can* be understood as meaningful within their particular

contexts and, as such, provide important functions for digital ethics. However, a critical and reflective approach to collectivising notions and ideas is important at all times to avoid negative downstream consequences for the actors, groups or collectives involved.

RELEVANT TECHNOLOGIES: WHICH CURRENT DEVELOPMENTS IMPACT DIGITAL ETHICS?

Along the lines of current trends in public debates around the world, many interviewees saw some **form of AI** – including neural networks, large language models (e.g. ChatGPT) and foundational models – which was introduced to the general market during our ongoing research phase, as particularly relevant technologies. AI was cited as being crucial given its potential to be all-encompassing with its potential for being entangled with many other existing technologies in unpredictable and unforeseeable ways. AI was also seen as fundamentally ethically challenging due to its innate lack of transparency. Some interviewees pointed out that we should refrain from calling AI “intelligent” as AI merely reproduces and recombines what is already there. Due to its self-developing inner complexity and ever-expanding nature, accountability or auditability were always going to be impossible, interviewees noted. They also stressed that part of poor accountability in AI also stems from the fact that its sources are not transparent. The products created by generative AI or the decisions made by AI are very often hard to explain and hence problematic. Similarly, generative AI raises questions about intellectual property and copyright. In a scientific context, interviewees noted, knowledge creation and reproduction processes as such will be altered through AI. Not all interviewees were against using AI, and many highlighted its positive potential and/or the areas in which it is already being used in largely positive ways, such as medical imaging diagnosis or in-vitro-fertilisation. Yet, most interviewees stressed that we, as a global society, need to engage in serious debates around liability issues arising from AI in order to find a common solution for potential harms.

The **Internet of Things (IoT)** was named as a particularly crucial technology for digital ethics for its potential to connect the online-world with the offline or non-digital world. The IoT holds potential for quickly and lastingly manipulating everything in the physical world connected through IoT objects, including energy sources, household appliances, smart entertainment, locks, fitness trackers and medical devices such as wearable diabetes technologies. Some interviewees stressed that although the IoT is a technology that connects the internet to physical objects, we should refrain from differentiating between the internet and other non-tangible technologies and the “real” world. On the one hand, the omnipresence of the internet

and other non-physical technologies in our everyday lives render them fully part of lived reality. On the other hand, seemingly “airy”, non-physical technologies always have a physical, energy-intensive counterpart, such as servers. This latter fact is obscured when we do not consider the “real”, hardware-driven aspect of the internet. Obviously, the internet itself was also named as a significant technology, having become so ubiquitous and such a routine infrastructure that it almost appears to be no technology at all. This, in turn, may cause grave ethical concerns as we may no longer be able to envision realities without it, keeping humanity in a relation of dependency, even if the negative effects were to outweigh the benefits.

Social media platforms were named as relevant for their potential benefits and dangers in terms of disseminating information to large audiences at a speed that is historically unparalleled. This provides people, interviewees noted, with platforms to spread questionable information and gather a following beyond traditional pathways to power. This may be utilised for societal manipulation and control. On the upside, social media could be used to deconstruct monopolies of power and to connect, with the COVID-19 pandemic being an obvious example. Simultaneously, social media platforms themselves are not neutral and can change in unpredictable and sometimes problematic ways. During the AGIDE research phase, Twitter changed ownership and was rebranded as “X”, which was mentioned by many interviewees as an example of a negative shift in power dynamics in social media platforms themselves. A strong regional and cultural difference was also mentioned in the interviews, with some countries still relying heavily on certain platforms which may have become almost redundant in others.

More hardware-based technologies, such as **(care) robots**, were also seen as ethically relevant. Robots have the potential to shift ideas about who is seen as an (non-human) actor in society and reconfigure interpersonal relationships and attitudes towards machines and their agency. Attitudes towards robots as part of the landscape of our everyday lives may show significant cultural variation, more so than may be the case with other technologies mentioned in the interviews. For example, as one interviewed expert noted, Shintoism’s animism approach allows for physical objects – not just humans – to be endowed with a spirit, arguably fostering relatability and relationships with robots. Robotics also carries meaning beyond the commonly debated ethics of care: as robots appear more human-like and are instilled with more complex software features, such as combining them with AI for communication, this may fundamentally alter the perception of a need for human connection and companionship, some experts warned.

Platform technologies perform numerous functions: they can offer products or services as a digital marketplace or take on the form of engagement platforms. Interviewees

stressed that all these aspects of platform technologies held the potent potential to match users with their (perceived) needs. However, this may also entail creating bubbles and mechanisms of exclusion. People may only be shown what algorithms “believe” they want to see. Platforms may fundamentally shape our worldviews by excluding other information and content. Interviewees considered a world seen increasingly through technology, especially platform technologies: platforms thus become the world we see – and depending on the algorithms that govern them, they show a particular type of differentiated, individualised, highly fragmented world. This may lead to a growing separation of groups and opinions in society, some interviewees cautioned. Moreover, platform technologies may (re) produce inequalities by only showing options according to the data they have collected and/or according to built-in bias. This could, for example, include a job platform not showing leadership positions to people it identified as “female”, given existing data, thereby confirming the historically low likelihood of people identifying as female being hired.

Facial recognition software was mentioned as often working together with other systems and, when utilised extensively, tracking movements on a global scale. Such software may, therefore, be used in systems of control and foster the loss of personal freedoms. At the moment, facial recognition, interviewees noted, works better for some groups than others – and is notoriously worse at correctly identifying people of colour (PoC) and known for inadequately identifying non-binary people. Interviewees stressed that, as a society, we face a situation in which it is unclear whether facial recognition software should be put to wide use at all. After all, in its current state, it is not even able to identify equally well all those who are involuntarily participating in its use.

Virtual reality was considered ethically relevant by many interviewees. As its expressions continue to be layered onto and intertwined with what we consider to be “real”, virtual reality blurs the boundaries between the technological and the physical. In that sense, virtual reality is similar to the IoT, but different in that it is not bound to objects already existing in the physical world. By enabling us to see and experience differently the world we know, virtual reality may also change *how* we experience it.

A more niche technology mentioned, but relevant for its large-scale impact, was **payment software**. Such software enables instant money transfers or payment using smartphones or other smart devices. Interviewees mentioned this as relevant for digital ethics given that payment software has revolutionised the workings of economies, shifting how money can be moved and how goods and services can be bought and sold on a fundamental level. Payment software, interviewees highlighted, may both empower workers and contribute to modes of exploitation in the marketplace.

DIGITAL ETHICS: MEDIATED BY OWNERSHIP, GOVERNANCE AND POWER?

A core finding of the AGIDE interviews is the embeddedness of theoretical and philosophical ethical principles in concrete structures of power and governance, with different manifestations depending on national and/or local contexts. Ethics, therefore, is not only embedded in laws but is executed in the form of daily practices, moulded by institutionalised power structures. Globally, decisions are being made with regard not only to the (non-)regulation of the digital, but also to the financing and non-financing of some digital products and networks over others. This creates manifold inequities and a multitude of directions in which digital regulations, and thus systems of use, will or will not develop. In what follows, we explore first how these questions of power emerge in regulatory structures, and then in more subtle and nuanced ways, such as through the ways in which the focus and priorities that drive technology development act to marginalise particular interests or populations.

Private Corporations vs. States: Regulatory Approaches for New Technologies

A key factor in this regard is the question of whether trust is placed either in corporate or state actors to regulate and govern digital technologies. Depending on their historical, cultural and political context, citizens’ willingness to share data with and place trust in either state or corporate actors varies greatly. For example, interviewees speaking on China noted that citizens felt it was in the interest of the common good to place trust in the state and have their data be used to regulate society, with industry having some regulatory influence. Meanwhile, interviewees speaking in a US context noted that the US had a history of little corporate regulation, so that citizens might have got used to freely sharing data with companies. Citizens of the EU, finally, were described in interviews as having got used to trusting state governments to keep them safe in their use of technology and as opting for a more cautionary approach in general. These differences, described in terms of collective populations, citizenship and attitudes, align roughly with the different regulatory approaches for new technologies, and the positions highlighted here represent the most differing views on the subject – with many other opinions in between and an array of differing cultural and political positions. One example that complicates the issue and relationships of trust are indigenous people and communities, many of whom harbour a deep-seated distrust of governments due to historic and current colonial practices and governments’ failures to uphold (treaty) promises.

At the time of writing, the US is proposing a model of corporate self-regulation with little federal intervention. China, on the other hand, is also concerned with technology

development, but focuses on state control more than the US has currently chosen to do. In contrast, the EU's AI Act takes an active regulatory, risk-based approach and tries to regulate new technologies based on the negative impact they could potentially have. The three contexts described here map onto key directions in which regulation of new technologies might be taken: *corporate self-regulation*, allowing for *state control* of technologies, and a *risk-based regulatory approach*. With the three regions mentioned (US, China, and EU) being prominent global players, many interviewees stated that other countries' and regions' approaches towards technology regulation could be loosely aligned with one of these three approaches. Each approach may be seen to have strengths and weaknesses, depending on the positions of the interviewees.

In the **corporate self-regulation model**, for example, we see that many large corporations have put ethical guidelines in place and have hired staff to check if their products follow these guidelines. Yet, some interviewees voiced the limitations of this self-regulatory approach given that corporate power over data and technologies is continuing to grow and is primarily driven by profit. Others believed that companies were often apt at self-regulating, as their utmost incentive was to gain and keep customer trust and they were therefore not incentivised to do anything to jeopardise that trust.

In the context of innovation, interviewees identified an upside of a **model of strong state control**, namely that having a player in place to oversee technology development, able to step in to regulate where need be, ensures that such development is not hampered. At the same time, such an approach ensures that corporate power and control does not become too strong. This approach, however, only works, interviewees cautioned, in the context of a functional, democratic government: if power shifted – as had happened in some of the countries interviewees were reporting from – having governmental control over technologies could quickly become dangerous for citizens.

The **risk-based model**, in turn, relies above all on a correct assessment of the risk at hand, which some interviewees were sceptical was currently always the case: after all, some technologies are increasingly all-encompassing, while others are, in fact, low-risk but falsely categorised and thus overregulated. Moreover, an approach where citizens are always asked to give consent to use websites and other appliances does not necessarily increase safety, but might simply place the (perceived) responsibility and control in the hands of individual users. In many cases, such users do not have the knowledge or power needed to implement change – making their consent virtually meaningless. Some interviewees stated that even more regulation is needed to ensure a “safe future” – with some going so far as to say that it might mean stopping the development of some technologies entirely until an accurate risk assessment

and action plan can be made. Others, in turn, identified a downside to this approach in its potential to cap innovation through overregulation and – sometimes unwarranted – fears. At the same time, the cautionary approach can never stop technology emergence wholesale and (overly) cautionary regulatory contexts would merely ensure that those jurisdictions are left behind economically and unable to actively shape the direction of technology development. Moreover, interviewees stressed ethical concerns over not using new technologies whose benefits might sometimes even be life-saving.

Other Stakeholders and Limitations of a State-Corporation Dichotomy

The three regulatory models mentioned in the interviews do provide a helpful heuristic spectrum for categorising and analysing regulatory approaches. As such, they could function as a means to open up debates around digital ethics regulation and governance. However, these models in no way represent an exhaustive list. Interviewees discussed several more nuanced approaches.

While governmental actors at state or regional level and corporations are not the only institutions regulating technologies, they are the ones that emerged most prominently in the interview data as key stakeholders. Nonetheless, other actors such as non-governmental organisations (NGOs), universities, regional and international interest groups, and citizen groups were also mentioned as having some power in regulating technology use and applying their ethical principles and priorities in so doing. Moreover, these groups execute regulatory power by influencing states and corporations through their behaviour and activism. Consequently, the portrayal of a state-corporation dichotomy in regulating technology, and thus inscribing digital ethical principles through regulatory processes alone, is never as clear-cut a matter as theoretical regulatory models may imply.

Political Systems and Control

Digital technologies have increasingly become an almost unparalleled lever of power for political control. While technologies' (ethical) influence almost always needs to be considered on a global scale, their impact can also be witnessed in local contexts. Digital tools, as many interviewees detailed, are used to gain or hold power in unprecedented ways, and often in “traditional” state contexts. Interviewees mentioned many instances of social media being used to shape and sway public opinion which had a significant impact on election outcomes. While “traditional” (print) media has been used to gain votes and shape opinions in the past, the speed and frequency of social media in relaying content has increased greatly.

While “traditional” media, interviewees noted, are often controlled by those already in possession of (some) power,

social media, including various messenger apps, allow for fringe groups or certain individuals to spread their message and quickly gain a following. Their messages are accessible – at least, in theory – to anyone with a smartphone. On the one hand, interviewees considered this a positive development, as traditional but publicly unfavourable power dynamics could be broken without large economic or political resources. On the other hand, the spread of so-called “fake news” and misinformation has reached almost uncontrollable levels, interviewees cautioned – most prominently visible during the COVID-19 pandemic. Spreading fake news through social media platforms has also become a tool exploited by existing governments or politicians in pre-election periods to retain or gain voters, for instance around the anti-vaccination movement in Brazil. The sheer speed of (false) information was considered the key ethical concern in this context and thus its ability to influence public opinion, even when information is removed by regulatory means. In turn, this might even come to strengthen conspiracy groups’ perception of being manipulated and controlled by higher powers.

This poses a multitude of ethical dilemmas. Fragmented public opinion may inhibit reaching common ground and common goals as part of democratic governance. Not all interviewees agreed that the concept of so-called “filter bubbles” (environments in which people are exposed only to opinions that conform to their existing beliefs) has been created through social media or is a new phenomenon. At the same time, many did state that it is now easier than ever to only consume news and information that aligns with one’s pre-existing notions and ideas about the world.

Apart from the power of social media to spread (mis) information, contemporary political control, interviewees noted, is also gained and exerted through significantly more sophisticated tools and tactics. With large amounts of data collected about citizens, they can be targeted directly not only through tailored advertisements, but also by being socially pressured into voting. Data may be and often are used to track individuals taking a critical stance towards governments, with interviewees giving examples of individuals suddenly charged with an (unrelated) crime to silence them based on digital data. Interviewees further noted the use of different kinds of so-called “spy software” produced by some states, thus leveraging a technological edge to gain power elsewhere. The strategic trade in spy software may also put the seller into a unique position to shape global power dynamics by way of a technological (weapons) advantage.

Lastly, corporate power and state power may be and are combined for interest groups to gain control. This, of course, calls into question any neat state-corporate divide. While country-level or regional governance may still play an important role, governments are, according to some interviewees, often so dependent on large technology

providers that those in control of technologies are increasingly powerful even within national contexts. Interviewees especially pointed to large-scale corporations, such as “The Big Five”: the most influential US-based technology companies, GAFAM (Google, Apple, Facebook, Amazon and Microsoft). Other similarly but more locally influential companies also play an important role in some contexts. Laws might be enforced to (try to) regulate large national and international technology corporations, yet some interviewees pointed to the impossibility of genuinely exerting state control over companies that are too influential at such globalised levels. While some interviewees mentioned the possibility of sanctioning companies – and named examples of countries or the EU as having successfully done so – the question remained what role corporations might play in (further) dismantling pre-existing global power structures and the role of the nation state in decades to come.

Risk and Risk Management

Outside of state and corporate control in the digital space, another key concern around digital governance noted by interviewees concerned different approaches to risk management and tolerance. Some interviewees stated that risk needs to be managed and anticipated as best as possible. In that sense, if a matter appears too risky after thorough assessment, “global society” should ideally not pursue that matter at all before appropriate risk mitigation. The sheer scale and wide-reaching impact of digitalisation make such a cautionary approach necessary, interviewees argued. Possible effects of risky systems “already out there”, they noted, are never fully reversible. That could mean overregulating and controlling systems that are by general accord considered to be too dangerous. A prominent example of that position mentioned by interviewees is the group of scientists and public figures speaking out against further AI development in March of 2023.

The opposing view espoused by other interviewees was that this type of risk-conscious or cautionary approach would inhibit innovation and its many potentially positive effects. Some interviewees also pointed out that a risk-conscious approach could, in any case, not be enforced globally. Therefore, it would de facto only prevent some regions from moving forward and keep them from innovation – inhibiting even positive ethical effects of innovation (see above). Interviewees argued for a risk-positive approach which might mitigate only selected risks, given the fact that not using certain digital solutions would come with its own dangers. Innovations and “new things”, they noted, often appear more acutely threatening than existing dangers. A commonly cited example here is the use of health data in a controlled, regulated manner to better plan for and regulate public health threats, such as the COVID-19 pandemic. While citizens, interviewees noted, were afraid of their data being used by the State, they simultaneously

ignored the threat posed by the pandemic if data was not used for their protection.

Global Governance: Diversity vs. Uniform Regulations

Thinking through the best kind of governance of digital spaces, many interviewees pointed out that most other countries follow one of the three core models symbolised and espoused (for the most part) by China, the EU and the US, as discussed above. Yet, they noted, some form of global regulation would also be needed, or at least a global commitment to agree on some common ground. Many aspects and infrastructures of the digital world, they argued, are global by design: for example, data for AI is generated continuously at an international level and many social media platforms operate on a global scale. Therefore, governing the digital space in a way that includes principles of digital ethics would ideally be done on that global scale.

For those interviewees who advocated for some version of global regulation, different approaches emerged: some proposed global discussion fora suited to agreeing on and prioritising ethical principles for the digital space, as is currently already being done by institutions such as the UN (see the discussion on the *UNESCO Recommendation* in Part A). This could then lead to a set of guiding principles, e.g. formulated in a digital ethics charter, which countries could implement in their own specific setting. A limitation of such an approach might be that, as discussed above, the same principle could mean different things in different cultural contexts and could thus be interpreted, enacted and take effect quite differently. Then again, this might well also be its positive feature, as it would allow for flexibility and cultural specificities of interpretation. Again, others argued that such openness to interpretation might render the charter meaningless in its effect.

An alternative approach to a charter of loose sets of guiding principles would be to reach international agreement on some hard laws. Such laws could then be enforced everywhere, as is already happening, at least in part, in the pursuit of cybercrimes. Here, the idea would include cross-national collaboration to ensure internet safety. However, interviewees cautioned, there are limits even now in how cybercrimes can be pursued across borders. Imposing ethical principles might pose even more of a challenge due to the different nature of penal versus ethics codification.

Yet another position offered by interviewees was that any expectations put on global governance are unrealistic. They cannot achieve their aims and should therefore not be pursued. Global governance, they surmised, is perhaps not even a desirable goal since it would take away from individualised solutions and regulatory efforts that make sense on a national or local scale.

DIGITAL EXCLUSION AND MARGINALISATION

As explored in the previous section, the governance of technologies has the potential to reinforce existing power structures and sometimes create new ones. The digital realm does not exist outside these structures of power. Inevitably, therefore, mechanisms of exclusion and/or marginalisation – be they unintentional or by design – exist in like manner.

As our world becomes increasingly digital, excluding some groups completely, or at least making the bar to entry for them too high to access digital technologies, might well be said to be equivalent to excluding people from society overall. Mechanisms of exclusion and marginalisation therefore pose immense ethical concerns. Digital exclusion, however, does not only take the form of completely barring individuals or groups from access. The digital space also entails many more subtle forms of exclusion and marginalisation which appear to allow the use of technologies in principle but render them de-facto inappropriate. Such de-facto barriers might be related to specific needs (relating, for example, to ways of accessing digital technologies) and/or to mechanisms of discrimination and bias detrimental to specific groups of users. At the same time, movements to create more universally usable technologies are active globally. They hold the potential to change digital systems, and thus society at large, in favour of more equitable outcomes.

Marginalisation as a Central Concern for Digital Ethics

Digital technologies are conceptualised, designed and built by a minority for a perceived majority which is thought to be similar, and to have similar needs, to those who design these technologies. Such thinking, most interviewees were clear, is ethically flawed. It leads to numerous inbuilt mechanisms of exclusion both in access to and in use of digital tools and applications.

Building technologies without having actual user groups in mind, interviewees noted, leads to false assumptions that those excluded are few and far between. However, those excluded – along with those “passively” affected by the existence of those technologies in their lives or “lifeworlds” – are in fact “marginalised masses”, as one interviewee summarised. They may even be a statistical majority of the (global) population. Yet, owing to mechanisms of structural inequality that cause them to lack the power to participate in, let alone sway, debates and design processes, excluded groups remain overlooked both within local contexts and on a global scale.

Excluding certain groups from digitalisation creates immense ethical challenges. Their concerns and their views on digital ethics are, by and large, unheard and ignored, creating digitised societies with the ethical principles of only a minority baked into systems that are often difficult to change once established. Mechanisms of marginalisation

as a concern for digital ethics may take place along the lines of, e.g., gender, location, social class, disability, ethnicity, language and levels of education, and very much at the variable intersections of these characteristics. They also have a temporal component, as they may change throughout life as people become more marginalised in old age.

By thinking through barriers to use, approaches of “universal design” (understood as a design approach that aims at enabling people of all backgrounds and abilities), as many interviewees noted, could ensure that digital technologies are and remain accessible and appropriate to be used by larger swathes of the (global) population. This may not only benefit marginalised groups, but may also lead to a more user-friendly design for all users.

Exclusionary Infrastructures

Besides design, interviewees pointed out another key area to foster inclusionary practices and counter exclusion/marginalisation in the digital space: that of infrastructure. In this, they included both capacity and infrastructure building and solutions that could slot into existing infrastructures.

Interviewees generally highlighted the global inequity in capacity building. Some areas of the world have remained critically underserved, leading some interviewees to suggest globally planned and financed investments to provide underserved locations with the infrastructure necessary to (fully) participate in digitalisation. Many digital solutions rely on high-speed internet, which is not available everywhere and can therefore render them virtually unusable. Furthermore, it should be noted that for large parts of the world digitalisation does not rely on computers, but rather on smartphones. Interviewees considered it essential to build more applications (apps) specifically with this circumstance in mind in order to ensure wider usability. Exploring local needs and knowledge, they pointed out, might also add valuable information about new or adapted technologies that best serve specific populations, thereby informing decision-making and processes of innovation. This might, interviewees concluded, entail investing in comparatively low-tech solutions which are easily accessible and usable.

Lastly, exclusionary infrastructures are often linked to the languages programmed into technologies. With English being the *lingua franca* of the digital space, other languages have been pushed to the sidelines. Some more marginalised languages, interviewees warned, run the risk of being virtually unused or lost in the digital space. At the same time, interviewees named specific, often not-for-profit initiatives set up to digitally capture local languages, preserve local languages in the digital space or even use digital tools to enable younger generations to learn them.

One of the main arguments cited by interviewees for investing in capacity building, infrastructure and universal (or inclusive) design relates to the fact that the digital

space relies on vast amounts of data produced by global communities. In turn, political, and especially economic, stakeholders stand to profit from that data financially or otherwise. Interviewees suggested that while individuals might not profit directly from their data being shared, they should at least be placed in a position to indirectly profit from new (often data-based) technologies by having access to and being enabled to use them for their benefit.

Experts, Lay People and Interdisciplinary Collaboration

A significant cause of inequalities in access to and in the design of digital tools and services was identified by many interviewees as resting with those who design and build digital spaces and with those who actively manage them. Their positionality, interviewees noted, is often that of a white, male, colonialising perspective. Such a hegemonic perspective harbours the danger of assuming itself to be the global norm and building systems that are self-referential. An apt metaphor employed by one interviewee was that of “10,000 PhDs”, referring to the assumed number of people holding a PhD degree in information sciences or information technology (IT). While that number in and of itself does not encompass all people involved in building digital spaces, it was meant to showcase the limited number of experts in the field not only in terms of numbers and in their elite positions but also in terms of the limited perspectives they bring to the table while designing technologies for the future.

Inequities in technology design processes become apparent when looking not only at the people who build them but also at the *places* where technologies are designed and with whose financial capital: that is, in the Global North, and specifically in the USA. Here, most investment capital and the predominant number of tech companies, such as Silicon Valley/US-based IT start-ups, receive billions in investments, while non-Western companies or projects outside of mainstream tech landscapes often struggle to receive significant funding.

A similar logic can be applied to education and employment systems that influence the creation of digital spaces, interviewees highlighted: when programmers do not find sufficient employment in non-Western countries, they leave, causing dynamics of a global “brain drain” in IT. Such dynamics further reproduce a system where local experts cannot create technologies with local needs and solutions in mind. One mitigating approach might be to invest globally in educating more people to code and design digital tools and infrastructures *en masse*. That alone, however, as interviewees cautioned, would not solve the issue. Shifts in educational systems would ideally need to go hand in hand with talent retention. Here, it should be noted, digitalisation itself provides a way forward: with many of the required skills for coding taught for little or no cost online, formal education might no longer be strictly required.

However, IT experts are not the only profession required in processes to create a more universal digital space: Interdisciplinary collaborations, interviewees emphasized, may promise to create fairer, more accessible technologies in an affordable manner. Such collaborations are often better suited to anticipating possible downstream pitfalls in designs that would have otherwise come up only at a much later point. Having interdisciplinary teams work out solutions together might also allow them to assess local and specific user needs better in order to build technologies and systems capable of solving real-life problems in both local and global contexts.

A similar debate among interviewees revolved around the question of whether members of the general public (future users) should be involved in building or, at the very least, in conceptualising digital technologies of the future – and if so, how exactly. The question of whether, how and to what extent lay people should be included and/or may be able to contribute to ethical decision-making around digitalisation was indeed at the centre of discussions around public involvement in digital ethics more generally. Opinions among interviewees varied greatly on this issue. One line of thought noted that it is impossible to opt out of digitalisation, and therefore lay people (qua users) should be increasingly educated about digital ethics and perhaps even be asked to assume responsibility by actively engaging in digital ethics debates. Given that the ethics of digitalisation concern them in day-to-day interactions both with technologies and in society, it is, interviewees noted, in people’s own best interests to think through the aspects of digital ethics meaningful to them. In contrast, another line of thinking argued that lay people could not be expected to educate themselves or even concern themselves with digital ethics debates; such a debate is an elite debate to begin with, and most digital ethics principles are set at too high a level to truly concern publics.

A common argument made by both sides outlined above was that lay people and experts alike often choose convenience of use and the upsides of using technologies over grievous concerns with digital ethics. This fact, indeed, makes it very easy for companies to bypass ethical standards and principles, both groups noted: if a product is convenient or attractive enough, a large majority will tend to use it, even when associated with ethically questionable standards.

Non-digital “Natives”

Many states and societies are actively moving towards increased digitalisation. This development includes state service provision, such as healthcare and social services, with the aim of employing more data in governmental decision-making. Consequently, interviewees warned, solutions for so-called “non-digital natives” are becoming increasingly relevant to avoid exclusionary states and spaces.

The term “(non)-digital native” – that is, people who literally do not feel “at home” or are not apt at using digital tools, services and technologies – was itself questioned by some interviewees. They noted that the term implies that an assumed “younger” generation can functionally use all online services with ease and by default, which is not actually the case. The term also does not address people who are excluded from digital services due to disabilities, not being able to read or understand written language, or who are not able to speak one of the required languages in highly digitalised states. Measures to “digitalise” citizens, if they are to be implemented, therefore have to address more than merely computer or smartphone skills, but also underlying assumptions and social makeup, interviewees cautioned. Speaking of a “digital gap” or a “digital divide” might be more accurate than to speak of (non-) digital “natives”, they noted. That term may still imply that the digital itself is the key barrier rather than the specific design of (state) infrastructures and services.

Some interviewees explained that the extensive need for outside help from relatives or friends required by many to access services is often overlooked in efforts to evaluate the percentage of citizens struggling to use digital services. When services depend on such outside help, they cannot really be considered accessible or easy to use. This also poses an ethical problem around independence and privacy rights. As societies move towards ever greater digitalisation, services must also be provided for those unable or unwilling to access digital services if states and non-state actors truly want to create non-exclusionary digital spaces. Some interviewees noted the example of countries such as Denmark, where public transport interval signs are no longer displayed at physical stops. Here, digitalisation is quickly becoming exclusionary not only in access to state services but also in terms of living in the public sphere more generally. Interviewees posed the question of how digitalised we would even want our physical world to become, at what cost and what would happen to those who could not or did not wish to join digital developments.

Finally, some problems framed by interviewees as issues of the “digital divide” turned out to be societal issues at large, expressed as, or perhaps distorted to appear as, an issue of failures in digital tools. For example, social services may be tied to a specific digital application and welfare entitlements might be withheld if people fail to use a provider app. On closer examination, however, it may well turn out that such failures are an issue of the welfare state as such, withholding entitlements on purpose or through more subtle mechanisms of exclusion. Such issues therefore point towards societal inequities more generally that can usefully be distinguished from their digital expressions.

TECHNOLOGICAL SOLUTIONS FOR DIGITAL ETHICS PROBLEMS

Given the complex web of positive and negative developments around digitalisation, interviewees suggested that some of the solutions for problems identified may well be solved by digitalisation itself: either by a better design process and/or by technological solutions that address digital ethics concerns directly.

Some interviewees stated that digital ethics principles and goals, as well as possible ethical challenges, should be considered from the start in design processes of products and services, e.g. by involving interdisciplinary experts in the process and using an ethics-by-design or resilience-by-design approach. Those approaches may, but need not, be more labour-intensive in the early stages of development but could potentially pay off later by creating systems with fewer malfunctions and making for a better, more universal user experience. Other interviewees disagreed, claiming that it is easier to build ethical codes or guidelines into already existing technologies – as a real-life test is the best way to uncover potential misfires and failures – and then to go back and fix them as needed.

Specific ethical concerns might also be addressed by digital solutions as such. A digital twin model, for instance, creates digital counterparts to (future) real-life systems or processes. Such models allow for simulations and testing of future products in a controlled environment. In terms of digital ethics, digital twinning allows for testing in a space without (negative) consequences for users, thereby, ideally, enabling problems to be identified and solved before a product is enacted in practice and thus protecting users from potential harms. Aggregated data, in turn, may solve privacy concerns by combining and averaging individual data into high-level data, e.g. for statistical use. As individually-identifiable data would no longer be required, fears around loss of privacy rights and data handled with malicious intent or malicious outcomes might be effectively mitigated. Aggregation, if done well, interviewees noted, could ensure data use without any loss of statistical power. Synthetic data also addresses privacy and data protection concerns by providing a model that does not require real-life data. Data is instead created artificially and algorithmically. Synthetic data may then be used as a stand-in for running dataset tests and machine learning programs.

Digital solutions, interviewees cautioned, cannot solve all digital ethics problems and they should not keep us from having larger debates about ethical principles. Yet, digital solutions must still be considered as practical tools to help mitigate specific ethical concerns.

KEY AREAS OF APPLIED DIGITAL ETHICS

While digital ethics plays a crucial role in all areas of digitalisation, interviewees highlighted a few central areas where digital ethics are either of particular import or where particularly impactful transformations had been witnessed that warrant closer examination. Such key areas are health, the economy and the environment, and are discussed next. It should be noted that this is a selection, not a complete list, based on cumulative findings across all interviews. This means, in turn, that some issues particularly relevant to certain groups are not represented here due to the applied methodological selection and the necessarily limited scope of this report. This in no way reflects their importance. For instance, for indigenous groups, a critical area is how digitalisation fares in protecting or revitalising (traditional) language and knowledge—alongside ethical considerations of doing so, e.g. considering issues of appropriate sovereignty, inappropriate commercialisation, cultural misappropriation and so on.

Digitalisation and Health

Health is an area that has benefited from digitalisation in numerous ways. Health and healthcare will potentially benefit on an ever-larger scale in the future. However, health is also an area that deals with especially sensitive data. With patient data being both some of the most coveted and most private data, it raises numerous ethical concerns.

Debates around digital ethics in health and healthcare are especially pronounced, not only because that health is a basic human right but also because health is also the right that forms the basis for all others. This fact was particularly notable when interviewees argued that health is a “special case”, to be considered even more carefully than others. Consequently, social injustices become especially apparent in the areas of health and healthcare, and the same holds true for “digital health”. At the same time, we suggest that considerations of digital ethics in the context of health offer a unique opportunity to find ethical solutions for digitalisation at large, using health as a case study in which the value of digital ethics is inherently apparent and particularly obvious.

Some of the current ethical challenges in digital health include the seeming contradiction of allowing individuals and groups *not* to share their data without consent, while also building datasets which are as inclusive as possible (e.g. for health AI). Groups who have been historically abused and marginalised in medical systems are now often understandably unwilling to participate in studies and/or share their data. This may, however, lead to further and future injustices in healthcare for those very same groups, given that they are rendered unrepresented or underrepresented in large datasets used for the development of better care solutions. Injustices of the past may here be

indirectly reproduced. Possible solutions, interviewees noted, might be to invest heavily in designing medical trials that strive to represent all population groups, and by actively reaching out to communities historically excluded and discriminated against.

Another way of achieving more equitable digital health, interviewees highlighted, would be to invest in comparatively low-tech versions of new technologies. For example, telemedicine tools could be built not only to connect high-level experts in already well-equipped healthcare contexts, but also in remote villages to connect to a regular physician or other healthcare worker in the next city for a first assessment. The same technologies can therefore be developed (further) with different levels of access and needs in mind. Creating low-tech versions would often entail solutions that run on smartphones rather than hospital equipment. It would also allow a larger part of the global population to benefit from new developments in digital medical care. Many interviewees also pointed to data leaks in healthcare in the news as illustrating how health data was currently not being protected as well as it could be. Technological solutions, such as aggregated data or digital twinning, may play an especially large role here. Such approaches could potentially be used to develop AI for healthcare without having to use individualised patient data.

Digital Economies

Digitalisation has impacted global and local economies in a multitude of ways, from fundamentally altering the way money can be moved in digital space and within seconds, to the way in which products and services can be accessed in a global marketplace. This has brought income opportunities to remote areas and to people unable to work in more traditional jobs: for example, data reviewers working from their homes and smartphones while also engaging in care duties. As digitalisation can also provide lower-threshold access to education and training, people have also used the internet to qualify for jobs previously unavailable to them. However, the digitalisation of economic systems has also raised many ethical concerns, which interviewees addressed at length.

Exploitation of labour is not a new phenomenon and indeed entrenched in extant economic systems. However, its regulation has become increasingly difficult. In digitised economies, many people may no longer be considered employees but are de-facto self-employed “gig workers” in the “gig economy” created alongside the emergence of labour platforms. Gig workers are not, in general, protected by the workers’ rights afforded to formal workers and are therefore at extremely high risk of exploitation through unsafe work environments, low pay, unstable work and more.

Work around the materialities of the digital itself, such as mining for rare metals or recycling digital devices, includes hazardous labour conditions that may be threatening to

long-term health and even be immediately life-threatening, interviewees highlighted. While such poor working conditions have become well-known over the past decade, no sustainable solutions have been achieved so far to ensure a higher level of international worker safety in the material support of digitalisation. Interviewees also listed positive examples of communities coming together to collectively advocate for gig workers’ protection: for example, after several food delivery service workers in China were harmed in traffic incidents, companies had to change their policies to reduce stress on workers, and accident rates were successfully reduced. Yet, no large-scale or encompassing protections have been achieved as yet.

While the (digital) gig economy is often heavily criticised for being exploitative, some interviewees added that not everyone wished to be employed in a traditional manner, and digitalisation opened up the possibility of more flexibility with working hours and labour choices. In contrast, specialised and rarer skills are particularly coveted, again possibly privileging the already-privileged. While digitalisation may take away workers’ traditional pathways to uniting against exploitative labour conditions, it may also hold the potential for new forms of collaboration and connection. Internationally, workers have connected in collectives, often organised via comparatively low-tech tools such as WhatsApp, using that platform to provide products or services in collaborating groups.

One last aspect to consider here is the way algorithms may impact labour: through ratings, algorithms make gig workers highly vulnerable to high-performance requirements and potential abuse. Algorithms may also be biased along the lines of existing parameters of discrimination. On the other hand, one positive example mentioned in interviews was the “she-taxi” initiative in India, a female-owned taxi collective, taking service only from women and children. Such algorithmic/platform-based tools have the potential to find new solutions in existing environments of discrimination and create safer work conditions, while also providing a safe and comfortable service for customers.

Digitalisation and the Environment

An often-overlooked aspect of digital ethics is the severe environmental impact of digitalisation on our planet. Interviewees warned specifically that those responsible for creating the greatest negative impact also take very little responsibility for their actions. According to some of the interviewees, it is easy to overlook how digitalisation negatively impacts the environment because of its seemingly and misleadingly “airy” quality. Such metaphoric lightness is further underlined by extant terminology, such as storing data “in the cloud”. The reality, interviewees noted, looks very different: large server farms store data, consuming vast quantities of energy and producing heat that requires cooling and thereby uses up even more energy.

The physical waste associated with digitalisation is highly toxic, seeping into and contaminating the air, water and ground of places largely located in non-Western countries. The environmental burden of digitalisation is thus distributed unevenly and unfairly across the globe. The so-called “planned obsolescence” of physical digital tools, paired with intensive marketing around “newness” and “replacements”, continuously create more waste.

Interviewees shared visions of AI being able to (help) solve global warming and other environmental concerns. In small contexts, digital solutions to environmental problems indeed seem promising, they noted. However, at the moment, digitalisation is firmly part of the problem posed by climate change and the challenges of the Anthropocene.

DIGITAL ETHICS AND DIGITAL FUTURES: CURRENT REALITIES, FUTURE IMAGINARIES

Finally, one major aspect of digital ethics concerned future developments in the digital space, exploring how digitalisation in its current process of development will shape us in the future and how humans are navigating the digital space as well as collective futures.

The “Digitised Human”: Digital Ethics of Changes in Human Behaviour and Condition

“Existing outside of the digital”, interviewees noted, has become a near-impossibility in many contexts: in many parts of the world, it is no longer possible to participate in education, work or social life without using the internet and other digital technologies. This may fundamentally alter how we interact not only with the technologies we use, but also with each other.

The image of the “digitised human” stimulated discussions: while the “digitised human” may not be better or worse than humans before the impacts of widespread digitalisation, they are still fundamentally different, interviewees noted. For example, staying in touch and interacting with each other has become increasingly easy through digital means, platform technologies and communication tools. The scope of that development was put into stark relief by the COVID-19 pandemic, where digital communication allowed people to continue work online, if their jobs permitted, and to stay in touch with loved ones despite lockdowns and in-person risks. At the same time, such opportunities cannot belie the loneliness felt during the pandemic in many parts and by many people of the world, and digitalisation may even have exacerbated this development. Some interviewees warned that the apparent opportunity to meet (new) people online at all times may create a collective mindset in which people become replaceable and lose dignity as individuals.

Another shift noted by interviewees was people’s attention span, which has been trained to decrease continuously by

faster-paced, ever more stimulating content. This can have downstream effects in how we learn, think and work – and not always in wholly positive ways. Some interviewees argued that digital tools have trained us, as humans, to think and process information in an increasingly machine-like way, which may pose a risk both for those who are able to keep up and for those who cannot or do not wish to do so.

Lastly, the digital tools we use have the potential to shape how we perceive the world: through “filter bubbles” (see above), we are shown content tailored to our algorithmically conceived interests. Based on algorithmic calculations, we are grouped with people seemingly similar to ourselves and shown content accordingly. While convenient in some instances, this black box ultimately carries the risk of creating and reinforcing who we are by showing us a limited selection of choices in the world – be that in the context of online shopping, job searches, dating matches, the next book we might like to read or, indeed, our understanding of current events in the world.

Human-Machine Relationships

As digitalisation increasingly seeps into every area of our lives, it is perhaps not surprising that our relationships with the machines around us are changing in like manner, interviewees explained. Here, digital ethics approaches can and do vary: Some interviewees, for example, spoke from parts of the world which consider (care) robots critically, framing them in terms of a society’s unwillingness or inability to provide human care. Interviewees from other contexts, in turn, saw robots as a helpful and friendly addition to people’s lives. Chat services might allow people to hold entire conversations with no one (human) on the other end. As AI become more skilled, such conversations might increasingly feel more realistic, entailing a risk of confusing or even misleading interactions in machine-human relationships. Some consumer or customer service chats already replace human service workers “successfully”, interviewees cautioned. While that is not an ethical concern per se, interviewees also discussed incidents of people preferring a romantic relationship with a robot – a phenomenon some considered increasingly more likely to occur.

Dystopian and Utopian Views of the Digital Future?

In the digital era, and especially in terms of its ethical underpinnings, humanity’s future appears at stake. Interviewees shared visions of a society so segmented we would never again leave our “filter bubble” (see above) to look at what lies beyond. Yet, not all views of the digital future were this dystopian.

Interviewees pointed out that our collective quality of life, and life expectancy in years, is higher now than ever before in history (if unequally distributed). While resources are also distributed unequally, collective wealth of money,

knowledge and opportunities, are today higher on average than ever before. Interviewees also engaged in utopian visions where digitalisation would be used to solve the world's greatest problems in the not-so-distant future – including, ideally, the problems caused by digitalisation itself.

Many interviewees stressed the importance of considering digital ethics and integrating them into this future – be that future utopian or dystopian. Overall, the metaphor most aptly describing the process envisioned by interviewees on the pathways to that future was that of “riding the wave” of digitalisation, rather than trying to “stop the ocean”.

THE AGIDE WORKSHOPS: ANALYSIS

INTRODUCTION

In the course of 2023, three hybrid (digital and in-person) workshops were held in Vienna, each lasting between one and two days. The first workshop took place at the Austrian Academy of Sciences on 17 and 18 April, with participation both on-site and remote. The second AGIDE workshop took place on 5 and 6 June in online format only, and the third workshop on 24 and 25 October, both in person at the Austrian Academy of Sciences and remotely. The workshops focused on three main questions:

1. What is your vision of a “good digital future” within your cultural context or region?
2. When you step out of the “bubble” of the expert community, what are the views of lay people you meet “outside”?
3. What is the most annoying cultural stereotype with regard to approaches to digitalisation? Why do you find it annoying and what would you change about that stereotype?

In cooperation with academies of sciences worldwide and based on interview recruitment, across all three workshops a total of 29 renowned international speakers were invited to explore these questions in line with the overall goals of the AGIDE project. The workshops were led by the Austria-based AGIDE project team and steering committee and organised around the above questions.

Preliminary Remarks

Speakers generally found it difficult to identify a distinctive “cultural context” to talk about or on, as most of them had lived or researched in numerous places and did not feel comfortable associating one particular context with their statements; similar sentiments were shared on the topic of felt “stereotypes”. The question about potential visions for a “good digital future” resonated most with the participants: the summary below will therefore focus on some

of the statements made in regard to this question which emerged as the most striking. Speakers situated their statements in their own perspective, experience and expertise, but common themes, approaches or strategies on the topic of digital ethics were deduced by content analysis for the purpose of this summary.

Finally, a note on procedural ethics: in the following we choose to identify the speakers and participants who consented to be named and associated with their statements; all other participants, in line with the AGIDE ethics protocol, were pseudonymised. Of course, due to the extensive amount of input, it was not possible to feature all speakers in this section. Thus, the following part represents a selection of speakers and their contributions.

EXPLORING A GOOD DIGITAL FUTURE FOR ALL

Because foresight is inherently limited given the complexity of sociotechnical systems, AGIDE's exploration of good digital futures does not attempt to correctly predict the coevolution of sociotechnical systems or reduce uncertainty (Floridi and Strait, 2020). Nevertheless, human activity “requires knowledge not only of what is most likely to happen, but also of what might possibly and desirably happen” (Urueña, 2019, p. 1). In this sense, AGIDE aims to stimulate debates that do not necessarily focus (only) on risk prevention, but rather on resources and opportunities for (digital) well-being and a good digital future. Moreover, envisioning alternative futures is also a useful tool for building resilience and informing decision-making (Sardar, 2010).

In addition, future scenarios “are typically defined as stories or future modal narratives” that wish to “develop an inclusive space for enhanced flexible decision-making processes” and “can be understood as a socio-epistemic practice” (Urueña, 2019, p. 2). Experts of course co-create these spaces, practices and narratives as advisors and prominent figures in public debates. Thus, „innovation as a future-creating activity cannot be detached from a careful study of the narratives it re-activates or brings to the forefront of creating meaning in society. As ordinary people [cf. lay people] and experts frame innovation in stories, tell and share stories, and make judgements of technological futures based on such stories, recurrent technological-cultural narratives structure their imaginaries” (Umbrello *et al.*, 2023, p. 5; Jasanoff and Kim, 2015). To this end, the AGIDE project aimed to use these workshops to create such inclusive spaces for a global debate and, above all, to make visible the narratives that will be touched upon in the section below and analysed in-depth in the next chapter (Part C) of the report.

THE AGIDE WORKSHOPS I, II & III: VISIONS FOR A GOOD DIGITAL FUTURE

Although three questions were asked in all three workshops, the first question about visions for a “good digital future” proved to be the most fruitful and significant, because it resonated most with participants, and is therefore discussed in detail below. It is important to note that for the purposes of this summary, the identification of common themes, approaches or strategies on digital ethics was derived through content analysis. Of course, it would have been possible to highlight other compositions and categories at different levels, as the data allows for different types of comparison (which can perhaps be explored more extensively in future research). However, these compositions seemed to be the most prominent and noteworthy.

It should also be noted that in the following section we have chosen to locate and identify the speakers in line with the situatedness approach that the report focuses on, in order to get a fuller picture of the differences between the many narratives of digital ethics. To this end, the speakers below agreed to be named and associated with their statements in accordance with the AGIDE ethics protocol.

WHAT THE GLOBAL NORTH HAS BEEN OVERLOOKING

A good digital future is about communities being in control of their future path and vision.

Te Taka Keegan, who is descended from the Waikato-Māniapoto, Ngāti Apakura, Te Whānau-ā-Karuai ki Ngāti Porou and Ngāti Whakaaue iwi in New Zealand, argues for Māori to control their collective data and digital systems as a means of achieving their digital sovereignty, cultural preservation and decolonisation. He noted that Māori should be in full control of their technologies, data and data infrastructures, because only through such collective ownership and sovereignty could long-term Māori-specific benefits be ensured. Other entities, including the government and big tech companies, do not have an inherent understanding or ability to see what is important to Māori and subsequently what digital strategies are most appropriate. Consequently, a common call is for Māori digital tools and technologies to be developed by Māori for Māori.

A good digital future is about global access to technologies and resources.

Teki Akuetteh, executive director of the African Digital Rights Hub, ICT lawyer based in Ghana and non-resident fellow of the Center for Global Development, emphasised the importance of equitable access to these technologies. Today, Africa still has one of the lowest levels of access to and use of innovation-enabling technologies from outside the African continent. Since the cost of innovation is really high, when new ideas arise, they are often sold to

multinational firms because they could not be developed further for the African market. This means, firstly, that technology from abroad (which could be used to support innovation in Africa) is too expensive and therefore inaccessible, and secondly, that innovative ideas developed in Africa end up outside Africa for the benefit of others. Moreover, (neo-)colonisation led to the development of diverse legal and cultural systems in Africa that exist in parallel today and often lead to tensions. Thus, challenges need to be solved in their own contexts.

A good digital future is about global equality and fairness.

Moreover, as Teki Akuetteh explains, Africa has become the “resource centre” of materials for the “developed world” and companies sell products back at a price Africans cannot afford. Global inequality in the context of digital ethics means that some people are mostly unaffected and unharmed by technology and may enjoy the benefits of innovation, while others are impacted in negative ways and excluded from the benefits. Lorena Jaume-Palasi, founder of the Ethical Tech Society and international advisor to the European Parliament’s STOA Panel (Panel for the Future of Science and Technology), also explains that people who enjoy those benefits are often ignorant of the material conditions and colonial history implicated in the development of the technologies they use. Countries in the Global South are predominantly struggling with the negative effects of technological development, such as water and energy shortages, waste and climate change impacts. In addition, inhumane working conditions in data labelling, correcting and cleaning, and in copper, cobalt and lithium mines for technology development are rarely seen or accounted for. Those very material conditions reflect global power asymmetries. Hence, if a good digital future means a good future *for all*, fairness and equity in the distribution of benefits and opportunities of technologies and innovation have to be assured.

A good digital future is about overcoming assumptions and cultural stereotypes.

Throughout the three workshops, speakers were asked about annoying cultural stereotypes when it comes to digitalisation in their cultural context(s). It was noticeable that speakers from the Global South often shared and faced similar problems or stereotypes, such as feeling underestimated, undervalued, excluded or neo-colonised (e.g. by big tech companies). Issues of access and security tended to be overarching themes. In addition, several speakers from the Global South saw young people in particular as having great influence on a (good) digital future. Yet, they noted, youths from the Global South are often conceived of only as consumers of Western ideas – not creators who want to use technology to improve their lives and empower themselves (e.g. Bhatia and Pathak-Shelat, 2023). The theme of

recognising often marginalised voices as creators, innovators and relevant actors was shared by many and identified as one of the key messages for the Global North.

INTERCONNECTEDNESS, ECOLOGY AND HARMONY

A good digital future is about finding similarities in differences within an epistemically just discourse.

Emma Ruttkamp-Bloem, head of the Department of Philosophy at the University of Pretoria and involved in many renowned ethics projects of organisations that advise on ethics both globally and in the (South) African context, urged the need to use culture “not as scissors but as a needle”, that is, as a tool of interpretation that could help to cultivate mutual understanding and bring humanity together. As such, international conventions are important because they support internationally recognised values insofar as they reflect respect for the intrinsic value of people in their diversity. In turn, using “culture as calculus” could assist in celebrating diversity, i.e. differentiating and distilling cultural values in order to meaningfully include, rather than exclude, them in international agreements. However, discourses need to be respectful, epistemically just, empowering, participatory, legitimate, inclusive and transparent (Ruttkamp-Bloem, 2023). In this context, a focus on the interconnectedness of all humans with each other is particularly important. For this, the African concept of Ubuntu (e.g. “*I am because we are*”) would be especially helpful as a guiding compass to anchor and manage a worldwide debate.

A good digital future follows an ecology-based approach with a focus on sustainability, symbiosis and harmony.

Yi Zeng, professor and director of the Brain-inspired Cognitive Intelligence Lab, and founding director of the International Research Center for AI Ethics and Governance, both at the Institute of Automation of the Chinese Academy of Sciences, brings in another perspective for thinking about interconnectedness and harmony. Coming from the field of computer sciences, Zeng argued for a harmonious coexistence of humans and AI in a sustainable, symbiotic society, following an ecology-based approach. He saw (embodied) AI technologies as future partners in society who would have abilities for compassion and morality through “brain-mind-inspired AI”. Therefore, AI must be driven to understand the principles of harmony and interconnectedness. China’s vision of a good digital future, he noted, includes national principles, ethical norms and governance following the primary principle of “living harmony”, with which he referred back to concepts in Confucianism, Taoism, Buddhism, Wa and Ubuntu.

A good digital future means finding consensus about our role in digital development and updating cultural concepts.

Shoko Suzuki, professor emeritus at Kyoto University, specialising in the philosophy of science and anthropology, looks at cultural concepts in a global context from another perspective: by extending the debate of interconnectedness to networks that also include non-living entities. As the principal researcher at RIKEN and an adviser to many councils in Japan and around the world, including the Information and Communications Policy Institute (Ministry of Internal Affairs and Communications, Japan), she spoke about the need to recognise that the wisdom of people and culture may lead to the right solutions to problems. To do this, societies have to reflect in order to understand themselves, their cultures and their role within ongoing digital transformations. It is therefore necessary to update cultural concepts for the digital future, which means translating cultural resources to a planetary scale (cf. Berberich, Nishida and Suzuki, 2020; Suzuki, 2023). For a good digital future, we would need to adopt new methodologies for our understanding of the world that go beyond dichotomies such as the “infosphere”, where everything is interconnected, constantly changing and forming a network that exists beyond the boundaries of the living and the non-living (cf. Suzuki 2024a; 2024b).

THE FUTURE OF WORK, YOUTH EMPOWERMENT AND FINDING A BALANCE

A good digital future demands addressing in tandem a fair and responsible future of work.

Payal Arora is a professor at Utrecht University, digital anthropologist and author with expertise in user experience and inclusive design in the Global South, with a particular focus on the Indian context and region (cf. Arora, 2019a). For her, both a good future and a good *digital* future depend on a shared goal for social and planetary well-being. Work is therefore not just what we do for one another or how we, as a people or society, innovate (cf. Bhatia, Pathak-Shelat and Arora, 2024). It is instead about ensuring dignity for all and nurturing our planet and society simultaneously while doing so. We can either be enslaved by the digital and get caught up in a “race to the bottom”, or if done right, we can build the right kinds of global guardrails and make the digital work for us. Moreover, she noted, the question “*A good digital future for whom?*” was essential for the global perspective (cf. Komarraju, Arora, and Raman, 2022; Bhatia, Arora, and Pathak-Shelat, 2021). Thus, a global society must focus on taking ownership of the global nature of work and the global value chain and build ethical standards that are also enforceable. This can be done through independent auditing bodies that can serve as watchdogs for mitigating digital harms and optimising joint flourishing. Additionally, Arora pointed out

that everyone who uses technology is also providing free labour for tech companies (e.g. by sharing data, building content). Consequently, in a good digital future, using technology should be recognised as work. This view, in turn, would create the need to redistribute value built by tech companies with everyone's data, as well as putting marginalised people at the centre of technology design and policy (cf. Arora, 2016; Arora, 2019b).

A good digital future is a good digital future for youth.

Invited speakers from Thailand, Ghana, South Africa, India and the Middle East agreed to a large extent that ethical questions about tech development were really questions about enabling young people to realise their aspirations for a good (digital) future and to use emerging technologies in innovative ways, e.g. for entrepreneurship and building communities. They pointed out that young people believe in globalisation, in solving key problems and in securing a better future for themselves through technology. Novel forms of transcultural and transnational citizenship practices are also emerging through discursive engagement, such as sharing information, raising awareness and mobilising collectives (cf. Pathak-Shelat and Bhatia, 2019). Access to technology, digital literacy and education were identified as key factors for youth empowerment. For example, Soraj Hongladarom, professor emeritus of Philosophy and research fellow at the Centre for Science, Technology and Society at Chulalongkorn University in Bangkok, studies the role of science and technology in the culture of developing countries, with much of his research focusing on digitalisation in Thailand. In his presentation, he explained how people in Thailand see technology primarily as a business opportunity and a way to improve their lives, with potential risks taking a back seat. For that reason, there is a rush to digitalise, and culture-specific technologies and businesses, such as an AI-based astrology reading service, are trending in Thailand (cf. Hongladarom, Joaquin, Joven and Frank, 2023). In addition, people in Thailand understand technology from a Buddhist and Confucian point of view, that is, as something that will inform them about ways in which they can realise their full potential.

A good digital future includes digital literacy, trust and finding balance.

Moreover, Nisreen Ameen, associate professor in Digital Marketing at Royal Holloway, University of London, who specialises in cross-national and cross-cultural research in emerging markets with a focus on the Middle East, also explained that younger generations are the most frequent users of technology, but are not educated in how to use it. Technology is seen primarily as something that can improve their lives economically and also as a vehicle for political change. At the same time, because a good

reputation is highly valued in the Middle East, people are very conscious of issues such as data rights and traceability. Therefore, the potential threat to values which are very important in these regions, such as heritage, family and reputation, can also be seen as a reason why people are sceptical about the use of technologies. For example, Imad Elhadj, professor of Electrical and Computer Engineering at the American University of Beirut (Lebanon), spoke about how difficult it was for the Lebanese Ministry of Health to provide government health apps during the COVID-19 pandemic, even when ethical guidelines were followed and transparent. Elhadj therefore argues that technological solutions must be home-grown as well as transparent to gain people's trust. Otherwise, there will be fear and resistance. Zabta Khan Shinwari, vice-chancellor of Kohat University of Science & Technology in Pakistan, adds that LMICs (low- and middle-income countries), including regions within the Middle East, want to preserve their culture and values, which are seen as potentially threatened by technology that predominantly comes from the West. This includes finding a balance between the opportunities that come with technologies from the West and the threat posed by their association with more individualist values, which are seen as being in tension with a more collectivist and hierarchical society.

DEMOCRACY, HUMAN RIGHTS AND THE WELFARE STATE

A good digital future is about democracy and human rights.

Within the discussion about the role of democracy, Anat Ben-David, associate professor at the Department of Sociology, Political Science and Communication at the Open University of Israel, argued that a good digital future is inherently linked to democracy and to a legal framework that protects human rights. She saw Israel as an example of a highly digitised society that has continuously put digital ethics to the side while pushing digital innovation and digital governance (see "Surveillance Innovation Complex" in Cohen, 2016). While innovation is accepted in all walks of life, the regulatory framework has not been updated in decades. What is therefore urgently needed is a cultural shift that could raise public awareness.

A good digital future is one in which technologies contribute to social and digital welfare.

Anne Kaun, who is a professor of Media and Communication Studies at Södertörn University in Stockholm (Sweden), has conducted surveys of Northern Europe, including Estonia, Germany and Sweden, exploring their citizens' relationship with technologies being used by the welfare state. The surveys showed significant differences in awareness, trust and perceived suitability of automated decision-making in public administration (Kaun, Larsson and

Masso, 2023). This reflects, she has argued elsewhere, “historical differences in welfare provision or so-called welfare regimes” (Lomborg, Kaun and Scott Hansen, 2023, p. 1) which go beyond individual variables such as gender, age and socio-economic background. The authors also identified processes of translation of guidelines and regulations at the EU level into local contexts. She argued that it is not true, as public discourse might suggest, that technologies are too difficult for citizens to understand. In her work she has found that people have very specific ideas when asked about concrete scenarios that are explained in a comprehensive way. Implementation of technologies therefore needs to account for such contextual differences and the need for comprehensive translations.

A good digital future will call for new professions and the setting of boundaries.

Paul Nemitz, who is principal advisor in the Directorate-General for Justice and Consumers of the European Commission and the former director responsible for fundamental rights and Union citizenship, as well as the former lead director for the reform of the EU data protection legislation, and who spoke first and foremost from his perspective of the German and European regions, argued that we need to develop new professions for a good digital future, such as a so-called “engineer for democracy”, who would be concerned not only with understanding technologies but also with the impact of those technologies on individuals and societies. Such impacts would go beyond biological effects but also affect our rights, thinking and freedom. Furthermore, as a society we are obliged to try to invest in long-term understanding of technology and its long-term impacts on humanity. It would also be important to recognise the potential and opportunities in shaping technologies and innovation through policy choices and democratic decision-making. This would also mean rejecting some future visions and technologies.

MOVE FAST AND BREAK THINGS: THE “FIRST TO MARKET” APPROACH TO INNOVATION

A good digital future is about informed government and constructive debate.

Contrary to perspectives from Europe, Alan Cohn, an American lawyer and former government official who served as assistant secretary for the United States Department of Homeland Security Office of Policy from 2012 to 2015, introduced the workshop participants to the US regulatory model, which generally favours innovation and industry, and is largely, if not solely, ruled by commercial interests, innovation, as well as “first to market” and “move fast and break things” principles. For a long time, Cohn noted, the US did not see the value of heavy regulation in the area of emerging technologies, and an

individual’s right to be free from government interference was generally enshrined in the constitution. Positive rights were rarely adopted, which is why debates about whether there even was a right to privacy were difficult to understand from outside perspectives. For Cohn, debates about the future have to be taken out of dystopian realms and back into the public sphere insofar as they would induce constructive public policy debates and are supported by knowledgeable governments.

A BRIEF WORD OF CONCLUSION

A good digital future is about embracing plurality and global participation.

The workshop speakers agreed that countries and regions are not homogeneous; rather, they are characterised by diversity, differences, contradictions, inequalities and polarities. And as Manisha Pathak-Shelat, a professor at MICA (Ahmedabad, India) and a specialist in digital cultures, social change and transcultural citizenship, explains, given the plurality and diversity of the challenges facing the world, it is difficult to find an approach to norms or laws that are universal enough and yet do not overwhelm local identities and ignore people in their differences, but do not leave too much room for interpretation. A “culturally sensitive” approach also depends on our definition of “culture”, which may be seen by some as static, but can also be understood as an evolving and dynamic process that could guide us into the future (cf. Pathak-Shelat and Bhatia, 2019). And, Pathak-Shelat adds, because traditionally marginalised communities are also marginalised in digital spaces that are defined not by a geographical space but by a global one, thinking about participation needs to go beyond a functional understanding of participation, especially in conversations about ethics. Thus, thinking about communities and individuals participating in the making of digital norms is crucial in order to ensure that those who have the potential to be affected by technology have a seat at the table. Moreover, as technology breaks down the notion of geography and community, global solidarity is also crucial (cf. Pathak-Shelat and Bhatia, 2019).

THE AGIDE USE CASE SCENARIOS: ANALYSIS

THE “USE CASE” SCENARIOS – INTRODUCTION

With the aim of not only highlighting differences but also focusing on common ground, AGIDE employed a scenario-based approach on the first day of Workshop III held on 24 and 25 October 2023 to deliberate on “individual values” and “shared principles”.

Drawing from the interview data and the presentations held in Workshops I and II, the Steering Committee pinpointed six key areas pertinent to potential futures of

global innovation: public space, social media and democracy, AI literacy, non-human agency, data communities and climate. Based on these identified areas, we intentionally composed provocative “scenario” descriptions of technological futures that were partially or fully hypothetical and indicated potentially positive or negative downstream implications to stimulate discussion among participants. This approach was conceived of as producing “use cases” which capture potential technology usage and its implications for individuals and or society. The use cases specifically contained contradictory consequences of technology use to show the range of possibilities and not to favour utopian or dystopian future scenarios.

We carried out three scenario-based sessions with members of the AGIDE working group, who were selected to encompass diverse cultural and scientific backgrounds and academic age, ensuring that each scenario was discussed by at least two groups. Four groups (consisting of four participants each) were physically present and one group (consisting of seven participants) participated remotely via Zoom, with discussions spanning 30 to 40 minutes. On an individual basis, participants were tasked in advance with reading the scenario and making notes on post-its with regard to:

1. the values they associated with the respective scenario,
2. potential factors in the scenarios contributing to discrimination, and
3. an assessment of the likelihood of that scenario unfolding in the future.

Within the follow-up group discussion, participants were asked to focus on (1) identifying “shared principles” for a good digital life regarding the scenario and (2) pinpointing research necessary to achieve that good digital life. We subsequently brought the insights from all group discussions back to the plenary of all participating working group members, requesting one spokesperson from each group to summarise in a single sentence their main discussion points.

Overall, the results of the scenario-based working group discussion were noteworthy: given our methodological approach, they showcased a collective ability in all groups to extract shared values and principles. Within the six scenarios (see details on each of the scenarios discussed below), the results were relatively homogeneous among the working group members, and even across the five different groups, a lot of similarities could be identified. However, across the six topics (i.e. the six different scenarios), we saw significant differences in which values and principles were associated and discussed. This underscored the way that distinctions tend to lie in *how* values and shared principles are put into situated practice.

Discussing needs for future research, the working groups came to similar conclusions regarding critical areas of

concern. Those included enhancing participation in decision-making processes and ensuring technology development proceeded in a well-informed and ethically sound manner. The working groups also emphasised the necessity of ongoing, iterative assessments of sociotechnical arrangements and their ethical implications. Such considerations demonstrated a clear commitment to ensuring that the intersection of technology and society could evolve in a manner upholding a “good life” while progressing digitalisation.

In the following, we present the summarised results for each scenario – the text in the grey box is a shortened description of the scenario description provided in the workshop.

PUBLIC SPACE

In this future, public spaces deal with conflicting realities. Integrated sensors and cameras heighten safety but spark privacy concerns. Holographic displays offer tailored information yet raise fears of manipulation. Augmented reality fosters cross-cultural ties but exposes societal rifts. Parks employ innovative systems for plant life, stirring debates on tech vs. nature. Art installations, though limited, challenge boundaries of expression. Facial recognition aids first responders but invites surveillance worries. Citizens shape their community through digital interfaces yet question online voting authenticity and inclusivity. Sustainable practices are promoted, but scepticism surrounds motives for smart waste management and renewable energy. This future is a battleground of harmony vs. control and progress vs. oppression, with humanity’s fate at stake.

In navigating the intersection of technology and data use in public spaces, the issue of dignity was at the forefront of discussions. This was independent of what constituted a public space, be they shopping centres, parks or political arenas. According to working group members, it is imperative to acknowledge the continuum between online and offline public realms, ensuring accessibility for all citizens, including those lacking online proficiency (as might be the case with people who are older or do not speak/read the language). Similarly, transparency emerged as a central concern, with questions surrounding who was in possession of knowledge and information. The “right to be forgotten” loomed equally large, yet the challenge of data erasure persisted, especially in the context of training sets for large AI models. A deeper understanding of privacy was considered vital, to include a sense of intimate connection that could still be violated, separate from any data protection issue. Tailored data storage and usage policies, group members noted, should align with the nature of the public space in question, recognising that shopping centres may require different surveillance retention periods than,

for instance, train stations. Age discrimination was also discussed as a structural issue, as younger generations do tend to be more digitally confident. The nature of the political system in place would also raise the stakes, given that non-democratic regimes could exploit surveillance in public spaces, and even democratic societies carry a potential risk of that happening. The extent of state control over citizens, data quality and material access were also discussed as factors influencing discriminatory data practices. Balancing the meaning of an individual's data to themselves against the relative importance of that data to others generally requires a more nuanced understanding of the issues at hand and how individuals might relate to their data.

Responsibility to ensure the rapid accessibility of information classified as “public” should also extend to all entities involved, including major corporations. The energy footprint of pervasive systems has to be weighed against their societal benefits. Acknowledging potential biases in AI models due to skewed training data is equally crucial for equitable treatment of all. Transparency in information sourcing and provenance is paramount. Respecting the choices of an individual for privacy, if so desired, would need to align with the much broader notion of “human rights online” (as was indicated on the post-it). Values like decency, dignity and equality in virtual worlds have to mirror how such values are upheld in physical spaces. Similarly, values such as control, accuracy (of data), physical safety and public order were discussed as particularly important. In turn, principles like transparency, oversight, accountability, freedom and risk-benefit analyses were deemed essential for ethical decision-making.

To move forward, the working groups identified areas requiring more and more comprehensive investigation, namely grassroots research, sector-specific impact assessments, practical operationalisation and involving citizens from various jurisdictions in shaping collective digital futures.

SOCIAL MEDIA

In this future, social media's impact on democracy is a mixed bag. Some see it as a tool of suppression, driven by algorithms favouring conformity and propaganda. Others value its inclusivity and the freedom of expression it can promote. Online discussions range from toxic echo chambers to enlightening debates. The authenticity of digital polls is debated, balancing claims of manipulation and citizen engagement. Social media straddles surveillance and connecting distant communities. Political campaigns range from misinformation-fuelled manipulation to genuine dialogue. This future embodies the dual potential of social media to challenge and fortify democratic processes.

In evaluating the intersection of social media use and democracy, working groups found the likelihood of their interdependence to be extremely high. They unanimously agreed that such a future is already upon us. Key values highlighted were democratic participation, conformity, inclusivity and freedom of speech. Open dialogue, cooperation, fair access and citizen participation were additionally considered as critical components in striking a delicate balance in social media use and democratic processes.

Potential factors contributing to discrimination and challenges discussed were: political views and identities, disparities in skills allowing for digital participation, and education and overall access to technology. Greed, accumulation of power in the few and manipulation of opinions were noted as potential mechanisms driving discriminatory practices and results. Additionally, concerns about power imbalances, surveillance, digital divides, fragmentation and monopolies emerged as significant considerations.

Capacity building for digital and political literacy was identified as crucial, requiring a global collaborative effort. Ensuring inclusiveness of engagement is pivotal and working groups called for the need to keep a balance between individual rights and public interests. Enabling social justice through mechanisms such as fair access, solidarity and promoting autonomy and reciprocity is imperative. One group succinctly summarised this issue by asserting that digital ethics has to prioritise global and interdisciplinary collaboration and aim to balance public interests and individual rights through participatory processes.

AI LITERACY

In this future, the impact of AI literacy is complex and debated. It is either a cornerstone of education, shaping curious minds and aiding ethical decision-making or an exclusive privilege, deepening societal divisions. AI-driven education unlocks potential, while ethical courses navigate moral complexities. Collaboration between AI and experts advances fields like healthcare, emphasising transparency. Society either harmoniously integrates AI for progress or sees it as a tool of oppression, reinforcing inequality and division.

It is noteworthy that groups members took varying views as to the likely realisation of the positive aspects of the scenario described above. One group deemed the scenario as overly optimistic, leading them to consider it highly improbable. Conversely, the other groups all expressed a high level of confidence in its likelihood. Despite this, the discussions pinpointed similar core values, issues and principles as fundamental.

Working group members agreed that the future of AI literacy is based on a foundation of computer, digital and mathematical literacy. Such literacy is not only crucial for

understanding the current state of AI but also for adapting to its ongoing evolution. Groups saw gaps in such literacy emerging as early as primary education. Similarly, “ethical AI” would require a more fundamental understanding of how human and social conduct would structure and transcend the specifics of each new technology. Achieving this would need sustained effort from all stakeholders over time. While “explainable AI” was discussed as an admirable goal, its practical implementation was considered challenging, as tighter constraints on development may inadvertently create incentives for actors to operate outside of those boundaries, potentially leading to even greater misconduct.

AI literacy, groups noted, has the potential to mitigate corruption. An AI-literate society could help promote a balanced approach to progress, ensuring accountability and safeguarding human autonomy. Potentially discriminatory factors underscoring AI literacy were discussed as multifaceted, encompassing issues such as defining “progress”, combating inequality, countering AI monopolies and addressing disparities in knowledge. Unequal access based on varying levels of literacy, manipulation tactics by powerful actors, algorithmic discrimination and (in-built) biases would all contribute to potential challenges. Additionally, the need to monitor potential shifts in societal structures, with the potential emergence of a powerful “AI class” alongside traditional power structures, was identified as crucial. Values like choice, freedom and transparency were highlighted as central to AI literacy efforts; similarly, flexibility, human dignity and acceptance of imperfections were considered to play pivotal roles. Accessibility, democratizing AI knowledge and ensuring fairness were seen as aspects that need to be addressed.

In a world where participation is increasingly facilitated through technology, ensuring general “technological literacy” and enabling universal access to technology were considered paramount. Shared principles for navigating AI literacy include protecting autonomy, and promoting human well-being, safety and public interests. Transparency, explainability, intelligibility, responsibility, accountability, inclusion, equality and sustainability were all mentioned as fundamental guiding principles for ensuring AI literacy.

NON-HUMAN AGENCY

In this future, the rise of non-human agents sparks both concern and promise. They lack moral judgment, potentially leading to oppression and manipulation and generating worries about exploitation, privacy, and abuse of rights. However, they can also be valuable contributors, enhancing efficiency and collaboration with humans. Advanced communication enables meaningful interactions among all entities.

Society integrates non-human agents into governance, research, and creativity, thus propelling progress. This future demands a careful balance between ethical considerations and harnessing the potential of non-human agents for human benefit.

The future prospect of humans coexisting with non-human (human-like) digital agents appeared highly likely to working group members to become reality within the next 15–20 years. To navigate such developments, the following core values were discussed: human dignity, privacy, fundamental rights, progress, innovation, autonomy, increased freedom, emotions, equity, accountability and the preservation of a distinct sense of “humanity”. Such values underscored the need for a balanced integration of technology that empowers individuals while safeguarding their inherent rights.

Human interests and well-being, the groups noted, have to take precedence over profit-driven motives and procedural interests. Such safeguards would ensure that the development and deployment of digital agents prioritised the “betterment” of humanity. To actualise this vision, rigorous research efforts would need to preserve fair and equitable use of technology and data. Such research would be the foundation for establishing ethical guidelines and governance frameworks that safeguard human interests.

Yet, multiple factors could pose challenges: The so-called “digital divide”, marked by discrepancies in access and capabilities, is a structural problem. Lack of accountability, coupled with built-in biases, could skew the impact of digital non-human agents. Commercial interests might also take precedence, potentially sidelining ethical considerations. In light of such challenges, the discussion of this scenario put strong emphasis on preserving human autonomy and understanding, with the aim of mitigating bias.

DATA COMMUNITIES

In this future, the role of communities whose data is being used in AI development (“data communities”) is a complex battleground. They can either actively shape algorithms with inclusivity and transparency, ensuring fairness, or be marginalised and exploited, leading to biased systems and eroding cultural diversity. The outcome depends on prioritising inclusive participation and transparency, and respecting privacy and cultural diversity.

In the context of the role of data communities in AI development, core values such as the commitment to data ethics, open data governance, data security and maintaining high data quality were discussed as critical. Moreover, working group members called for a shared dedication to utilising data for the “social good”, fostering collaboration,

preserving privacy and upholding transparency and access. Additionally, education and training, diversity, sustainability, equality, accountability and fairness would play pivotal roles in ensuring that data communities thrive and do so equally. Such values would collectively aim towards achieving collective prosperity, well-being and the preservation of cultural diversity and identity.

Yet, several factors relevant to discrimination may influence the experience of these data communities. Availability of resources, awareness, potential misuse of technology, monopolistic practices and power imbalances were discussed across all working groups. Avoiding marginalisation, the ethical use of resources, empowerment and the concept of shared responsibility all have to be considered. The relationship between government entities and private corporations would further impact such data communities. To ensure the success and equitable growth of empowered data communities, it is imperative, according to working group members, to emphasise transparency, accountability and privacy. Equitable resource allocation and inclusivity have to be prioritised. By enacting those values and addressing factors, a future where data communities serve as catalysts for societal transformation is indeed imaginable.

CLIMATE

In this future, digitalisation's impact on climate and sustainability is a double-edged sword. It can either empower eco-conscious choices through smart tech and connectivity or be wielded to exploit resources, exacerbating environmental crises. Smart grids and AI-driven models can reduce emissions or become tools of control and misinformation. Virtual collaboration may curb travel pollution or prove inadequate. Renewable energy adoption can thrive or be hindered. This future's outcome depends on responsible governance and ethical practices.

All working group members deemed such a scenario highly likely, particularly emphasising its positive aspects. They deliberated extensively on core values like equality, fairness, resource distribution, allocation, access, inclusivity, sustainability and diversity. Additionally, they highlighted important principles such as safeguarding planetary health, preserving resources for future generations, and promoting sustainable development.

One group acknowledged that the likelihood of positive aspects of this scenario dominating hinges on the interplay between corporate entities, with a vital role reserved for civil society and the essential need to assess comprehensively AI's environmental impact and explore diverse interventions to heighten environmental consciousness. Notably, the potential drawbacks of immersive technologies were underscored and concern expressed that "gamification" should not take precedence as the primary method of teaching about climate. Thorough assessment and due diligence in evaluating technology's impact on sustainability were stressed as prerequisites; virtual and in-person interactions have to be balanced.

Potentially discriminatory factors were thoroughly discussed, including issues of unequal control, extractive practices, recklessness (in climate-related actions), and hypocrisy, as well as unequal resource distribution, manipulation, monopoly, surveillance and the imminent threat of climate collapse. Moreover, the discussion touched on the challenges of inadequate infrastructure in certain world regions and the potential inaccuracies and accessibility issues surrounding climate metrics. The impact of AI on sustainability is contingent on societal choices, groups noted, necessitating rigorous research on the environmental output of AI and a comprehensive understanding of the potential risks associated with augmented reality technology.

PART C: MAPPING DIGITAL ETHICS – THE NARRATIVES

SYNTHESISING THE EMPIRICAL FINDINGS

While the original aim and design of the AGIDE project was to uncover (cultural and other) differences in values and value prioritisation, the data analysis revealed that at a basic level, values are widely shared and prioritised more or less equally across the world (see Part B). However, through the systematic analysis of all the empirical data collected, i.e. the international expert interviews, the two thematic input workshops and the use case scenario analysis, it was possible to uncover another level of insight, namely differences in the narratives of digital ethics. For this reason, in this chapter AGIDE presents the differences in the narratives of digital ethics according to different regions, countries, communities or groups.

Given the wealth of data generated in the empirical parts of the project, the results could have been synthesised in many ways. For a better overview, we have chosen to visualise the findings in a “matrix of digital ethics narratives”, which presents eight key dimensions that are relevant to the analysis of these (regionally) situated narratives. For this purpose, the qualitative data was coded according to inductive research principles, then abstracted and grouped, preserving the relationships between the identified themes.

MAPPING DIGITAL ETHICS – THE NARRATIVES

There are many similarities between this part on narratives and the work of anthropologists on narratives, first and foremost in the fact that both have the same aim: the study of “narrative practices in different parts of the world and the cultural and situational particularities of their emergence and character” by investigating “how narratives help individuals and groups to make sense of experience; how narratives contribute to socialisation into group practices, norms or moral values; and how they help to transmit ideologies, theories or imaginaries” (Götsch and Palmberger 2022, p. 2).

Moreover, in a similar way to anthropologists, AGIDE understands narratives as stories or arguments that are *told repeatedly and consist of a series of events* that are selected

and arranged in a particular order, often including, among other things, *central characters* (protagonists, antagonists), *a conflict and a plot* (e.g. Götsch and Palmberger, 2022; Bamberg, 2012). In addition, narratives are highly sensitive to context, history, power relations and temporality (which is also reflected by AGIDE’s focus on the “situatedness approach”; Haraway, 1988). In this sense, narratives are not understood as an objective reflection of life, but as a social construction (e.g. Spector-Mersel, 2010, p. 208); a social practice (e.g. De Fina and Georgakopoulou, 2008) and a way of meaning-making that links experiences of self and society in a cohesive form (cf. “subjective-in-between”; Arendt, 1958, p. 180; White, 1987).

Since this “group effort of narratively reconstructing a shared experience” (Götsch and Palmberger 2020, 15) is society’s “main way of making sense of things” (Culler, 2011, p. 82), narratives that are shared by a wider group and/or promoted by influential actors, can become dominant in a particular social setting and powerful drivers of collective behaviour (e.g. Bamberg and Georgakopoulou, 2008; De Fina and Georgakopoulou, 2008; Rhodes and Brown, 2005; Booth *et al.*, 2009; Wittmayer *et al.*, 2019). Also, stories that align with dominant narratives are often reinforced and, if sustained over time, become “grand stories” (e.g. Le Roux and Oyedemi, 2022; McLean and Syed, 2015; Wodak, Reisigl and de Cillia, 2022; Groth, 2019), which are “stories common to the groups we belong to [that] we create our familial, organizational, community and national identities” (Spector-Mersel 2010, p. 208). Further, grand stories relate to ethics because they give meaning to events and (morally) guide populations as to “what ‘worthy’ life is, what we should aspire to and what we should avoid, what is good and what is evil” (Spector-Mersel, 2010, p. 208).

However, narrative practices do not only constitute the meaning of communities (cf. “imagined communities”; e.g. Anderson, 1983), but also of (cultural and regional) places (cf. “placemaking”; e.g. Palmberger, 2022; cf. “digital placemaking”; e.g. Haleboua and Polson, 2021). The social production of places, communities and cultural beliefs is also reflected in the abstract narrative models in Part C. Thus, even though certain regions and places are partly referred to geographically in this chapter (for lack of other ways of

describing them), it has to be kept in mind that this section is primarily about the social co-construction of narratives, identities and places, which were derived and “co-created” from the AGIDE interviews and workshops.

AN INITIAL MATRIX FOR ETHICAL NARRATIVES

The matrix, seen below, is but one way of structuring our empirical findings and key topics. Alternatively, we could also have focused on specific concerns that were often explicitly mentioned (such as “privacy”, “democracy” or fear of discrimination). The matrix, however, highlights patterned observations drawn from a large and diverse set of empirical evidence and analyses, i.e. the interviews and workshops detailed in Part B. As such, it provides a means to put on one plane ideas and notions around digital ethics which emerge in situated interaction with local contexts as well as globalised attitudes to digitalisation, innovation

and the social “good”. When put together, this results in characteristic patterns of stories, reflecting certain characteristic *ethical narratives*.

It is important to note that the matrix does not express real-world narratives so much as a systematised and grouped means of giving them a shared frame – and it necessarily generalises. Narratives are also not necessarily consistent, i.e. stories told by people in one situation or sector (e.g. social media and democracy) are not necessarily aligned with stories told in other situations or sectors (e.g. social media and hate speech), and even less so are stories told (e.g. concerns about privacy) always aligned with what people actually do (e.g. click consent buttons).

Below we present the eight key dimensions and the three-fold spectrum with which they are associated. They represent, respectively, key overarching themes and their different varieties or characteristic manifestations (similar to “archetypes”).

	Key dimensions	“Building blocks” of digital ethics narratives		
Fundamental ethical assumptions	Notion of “the good”:	harmony/virtue	deontological	consequentialist
	Notion of “fairness”:	role adequacy	material equality	formal equality
The protagonist and its role	Primary reference point of narrative (“the protagonist”):	ecology	community/society	individual(s)
	Position of reference point vis-à-vis technology:	beneficiary/convenience	victim/risk	actor/opportunities
The conflict and the antagonist	Primary ethical concern:	prosperity/marginalisation	safety/harm	autonomy/lack of freedom
	Actors potentially posing a threat to this concern:	government(s)	tech industry	(other) users
The wider plot (what should happen)	Governance tool to address ethical concerns:	culture/education	law/regulation	technology/economy
	Factor that gives legitimacy to governance solutions:	organic evolution	determination by the able	self-determination

Figure 2: Matrix of Digital Ethics Narratives

The three columns do not stand for particular regions of the world as such, nor for any specific communities. The matrix refers to a **characteristic emphasis put on a particular aspect in each of the eight dimensions** and lists the aspects in one of three columns. This, however, does not mean that a person who puts most emphasis on the aspect in the left-hand column in one dimension will also put the most emphasis on aspects listed in that column in other dimensions, i.e., the contents of the columns follow a

content-related rationale, but cannot be clearly delineated; a combination of all factors is conceivable in principle. On the contrary, it is the emerging characteristic pattern of emphases in the various dimensions that makes the story, i.e. forms a particular ethical narrative or pattern.

There is no single answer as to what produces the differences in narratives, and analysing the underlying factors would be beyond the scope of the AGIDE project at the time of writing this report. The data does suggest that there

is a range of different factors and that the intersectionality of these factors influences the choice of approaches and people's experiences. At a macro-level, these factors involve the political system, religion, economic system, geo-political position (both historical and contemporary), alignment with particular discourses on technology and globalisation, governance of the digital environment, and more. At an individual level, it could be factors such as age, experience with technology, gender, education, income, class, cultural background and similar that play a role.

The following definitions and explanations should help the reader understand the meaning behind the terms used in the matrix, based on the empirical findings and analyses of the AGIDE project.

EIGHT KEY DIMENSIONS OF DIGITAL ETHICS NARRATIVES

FUNDAMENTAL ETHICAL ASSUMPTIONS

The first set of parameters, i.e. the notion of “the good” and the notion of “fairness”, is of a more general nature: not necessarily specific to views with regard to digital technologies in particular but of essential underlying importance. This set of parameters forms the background to stories told by AGIDE participants. They were necessarily made explicit in the contributions to the first two AGIDE workshops. In the interviews, they were sometimes explicitly addressed, but more often, they figured as implicit assumptions.

1. Dimension: Notion of “the good”

Harmony/virtue: includes approaches to ethics that take as central some concept of harmony or virtue, i.e. a morally good disposition to think, feel and act well within a given framework. Note that this includes a wide variety of notions that have been prevalent in very different settings, from Confucianism to ancient Greek philosophy to Ubuntu.

Deontological: refers to ethical theories according to which the morality of an action should be based on whether that action is intrinsically right or wrong according to a set of normative standards (from whatever source). In “Western” discourse, this set of theories is usually referred to as “deontology” or more specifically as “duty ethics”.

Consequentialist: refers to ethical theories in which the consequences of a behaviour are the ultimate basis for judging the rightness or wrongness of that behaviour. This set of theories is usually referred to as “consequentialism”, with more specific subtypes such as “utilitarianism”.

2. Dimension: Notion of “fairness”

Role adequacy: refers to ethical approaches that stress the ideal of role-adequate behaviour, i.e. geometrical equality

(proportionality), often based on the assumption that different roles have been assigned by powers outside human society, such as God(s), nature, fate or the spiritual world. Note that this may lead to attitudes qualified as racist or sexist.

Material equality: refers to ethical approaches that stress the ideal of equality of results or at least opportunities, trying to adjust unequal starting positions by way of affirmative action or removal of structural barriers.

Formal equality: refers to ethical approaches that stress the ideal of equal treatment, accepting that unequal starting positions may lead to unequal opportunities and therefore results.

THE “PROTAGONIST” AND ITS ROLE

The second set of parameters begins with the primary point of reference or concern which or who constitutes the “protagonist” of the story told. This “protagonist” may also be the preferred protagonist in contexts other than digital ethics. However, the matrix specifically addresses the position of the primary point of reference vis-à-vis technology because this is a key element of any digital ethics narrative.

3. Dimension: Notion of “the primary reference point of narrative (‘the protagonist’)”

Ecology: refers to ethical approaches that tend to downplay a categorical difference between humans, animals and plants and the inanimate world, and take ecosystems as a whole as the primary point of reference (but not necessarily with a particular concern for environmental issues, such as climate change). Note that this is simply an extension of the notion of “community” or “society”, i.e. concern for the animate or inanimate world often extends only to the environment that serves a particular community or society.

Community/society: refers to ethical approaches that draw a clear line between human and non-human and that focus on human societies and their collective well-being (usually accepting, or even insisting on, corresponding duties of solidarity on the part of the individual).

Individual(s): refers to ethical approaches that focus on (human) individuals and their rights (often tacitly accepting or even insisting on corresponding duties of the community). Note that this is not incompatible with an emphasis on particular (e.g. marginalised) communities, as long as concern for these communities flows from concern for the individuals within them.

4. Dimension: Position of primary reference point vis-à-vis technology

Beneficiary/convenience: refers to attitudes that understand digital technologies as something that mainly benefits their primary reference point (e.g. the community or the individual). However, the reference point is largely

envisioned as a passive user. Note that this approach includes attitudes that appear to make “unreflective” or “uncritical” use of new technologies, focusing on convenience or on necessity (e.g. because opting out of using technology is next to impossible).

Victim/risk: refers to views that see digital technologies as something potentially risky or disadvantageous for their primary point of reference (e.g. the community or the individual), creating a need for protection. The risks or disadvantages perceived can vary and often include exploitation, manipulation, discrimination and/or marginalisation.

Actor/opportunities: refers to attitudes that see digital technologies as something that their primary reference point (e.g. the community or the individual) can actively use to achieve their goals. The main difference with the attitudes described under “beneficiary” is that “actor” implies a much higher degree of reflection and a sense of agency, that is, of being in control of developments and actively embracing opportunities.

THE “CONFLICT” AND THE “ANTAGONIST”

The third set of parameters is about the primarily perceived ethical concern (positive or negative) and the actor that may pose the most significant threat to this concern. In the story told, this creates the “conflict”, in which the “protagonist” is involved, along with a potential “antagonist”.

5. Dimension: Ethical concern

Prosperity/marginalisation: refers to attitudes that see the primary concern as prosperity or well-being (usually in a somewhat material sense, and if immaterial, more distant from notions such as autonomy). This is not incompatible with concern about harm, but the emphasis is more on fear of potential marginalisation (e.g. exclusion due to a digital divide in the population) or exploitation (e.g. one’s data being used for profit).

Security/harm: refers to attitudes where the primary concern is to be safe from harm. This is not incompatible with being concerned about exploitation or manipulation, but the emphasis is more on the fear of loss (e.g. through an accident, an adverse decision, etc.).

Autonomy/lack of freedom: refers to attitudes in which the primary concern is being able to make decisions for oneself. This ability may be impaired or even absent, in particular where one is subject to excessive surveillance (cf. the notion of informational self-determination) or manipulation. Note that one can be concerned about autonomy even when surveillance or manipulation does not lead to further downstream consequences or harm.

6. Dimension: Actors potentially posing a threat to this concern

Government(s): refers to attitudes that see the main problem as lying in the use of technology by the government (in

a domestic setting, e.g. for surveillance purposes) or by other governments (in an international setting, e.g. for espionage or warfare, or simply for gaining more economic power).

Technology industry: refers to attitudes that see the (potential) behaviour, increase in power, etc. of the technology industry as the main problem. Note that this also refers to attitudes that see technology itself (possibly even in a “personified” role) as the main problem.

(Other) users: refers to attitudes that see the main problem as the (mis)use of digital technologies by malicious actors. Insofar as these malicious actors are private actors (e.g. abusive business practices, cybercrime) they can be influenced or facilitated by state actors (e.g. troll attacks as part of hybrid warfare), in which case it may be difficult to draw a line between malicious users and “government(s)”.

THE WIDER “PLOT” (WHAT SHOULD HAPPEN)

Digital ethics narratives very often result in calls or recommendations for action, i.e. who should react in what way in order to resolve the “conflict” in which the “protagonist” is involved. The fourth set of parameters therefore addresses preferred governance tools for managing perceived ethical concerns, and factors that give them legitimacy.

7. Dimension: Governance tool

Culture/education: refers to approaches that see social norms, as they have been shaped over time by culture, religious beliefs and philosophies, as well as education, as the preferred governance tool (often downplaying the role of law, but not necessarily denying its necessity).

Law/regulation: refers to approaches that stress the role played by the legal system, and of clear rights and obligations, enforceable by a polity. Normally, this means action being taken by the government/legislature, or several governments (in the case of international law).

Technology/economy: refers to approaches characterised by a high degree of trust in the self-regulating forces of technology itself and the interplay with markets to shape the (digital) economy according to societal needs. Normally, this means action taken by (often big) companies.

8. Dimension: Factor that gives legitimacy to solutions

Organic evolution: refers to approaches that see legitimacy coming from the dynamic nature of the world and human societies and trust in the invisible forces at play, which tend to downplay the need for or desirability of intervention and openly embrace change.

Determination by the able: refers to approaches that see legitimacy in the intervention by actors (often the government or religious leaders) who are able to distinguish the “good” from the “bad” or the “right” from the “wrong”. In Western discourse, many of these approaches are considered “paternalistic”.

Self-determination: refers to approaches that see legitimacy mainly in self-determination (mainly of the primary reference point, e.g. the community or the individual). Note that self-determination can mean different things depending on what the primary reference point is.

CHARACTERISTIC NARRATIVES OF DIGITAL ETHICS – THE MACROPERSPECTIVE

From an (oversimplified) macro perspective, looking for differences rather than similarities in our analytical matrix approach allows for several characteristic patterns of “ethical narratives” to emerge. These narrative patterns are not necessarily characteristic of a particular region, cultural or other kind of context, but reflect the very particular “situatedness” of the individuals contributing to AGIDE. No one factor – be that religious, cultural or educational – determines these patterns of ethical narratives. Yet, certain characteristic patterns seem to be dominant in certain regions or countries, linked to different ontological, legal, religious, political or cultural traditions. Sometimes they represent a dominant narrative around certain regions or countries in that they would, if retold in public, likely meet with large-scale approval or understanding. It should be noted that analysing shifts and developments in narrative patterns can be even more interesting than the narrative patterns themselves, underlining the fluid nature of such patterns. In the following section, we detail these narrative patterns and analyse them further according to the eight dimensions and associated attitudes identified above. In so doing, we not only show the usability of the matrix as a tool, but also how this method in and of itself can be usefully adapted to different contexts for telling meaningful, situated stories on digital ethics within a shared, but not self-same, frame.

“COLONIALITY”-TYPE NARRATIVE PATTERNS

One characteristic pattern, which emerged in several contributions from participants who spoke about their experiences in and with regard to the Global South, was strongly associated with related (joint) experiences of colonialism. The starting point of such narratives was usually in a notion of the “good” that focused on different versions of “harmony” and a notion of fairness that emphasises “role adequacy”, although there were also contributions that seemed to focus more on compliance with standards. As to the primary reference point, many contributions showed a focus on ecology as a broader reference point (with the community or society still at the core of any given ecosystem, often understood in ethnic terms). However, given that this “Coloniality”-type pattern was shared across different regions, there were also variants that took compliance with standards (such as fundamental rights) and material equality as a starting point, which in turn, did not seem to indicate strong ethical concerns beyond human community/society.

Common to most approaches reflecting this pattern was a strong sense that the community in question was or might (again) become marginalised and exploited and, in that sense, a potential “victim” of digital technologies and those who develop and deploy them. Approaches reflecting this pattern often referred to stakeholders representing other world regions, such as the governments and industries belonging to those world regions (often phrased in terms of “digital colonialism”).

Among the most common and prominent ethical concerns were therefore those of community autonomy (self-determination). In terms of the actors trusted to address that concern, local (small) businesses and initiatives seemed to play the most significant role, often with an emphasis on developing local technical solutions.

Fundamental ethical assumptions	harmony/virtue	deontological	
	role-adequacy	material equality	
The protagonist and its role vis-à-vis technology		community/society	
		victim/risk	
The conflict and the antagonist	prosperity/marginalisation		
	government(s)		
The wider plot (what should happen in terms of governance)			technology/economy
			self-determination

Figure 3: “Coloniality”-type pattern

We have also observed a second narrative following the same schematic type, for example from contributors from the African continent, which puts the emphasis on a plea for (international) law and regulation, to be put in place

by governments and international organisations. Associated with that was the hope for such regulations to provide adequate safeguards against exploitation and new forms of (neo-)colonialism in the digital age.

Fundamental ethical assumptions	harmony/virtue	deontological	
	role-adequacy	material equality	
The protagonist and its role vis-à-vis technology		community/society	
		victim/risk	
The conflict and the antagonist	prosperity/marginalisation		
	government(s)		
The wider plot (what should happen in terms of governance)		law/regulation	
			self-determination

Figure 4: Variant of the “Coloniality”-type pattern

“BENEFICIARY”-TYPE NARRATIVE PATTERNS

A very different group of narrative patterns is what may be called the “Beneficiary”-type narrative patterns. This type of pattern was dominant in contributions from a broad range of regions, many of which find themselves equally associated with the Global South or perceived as emerging economies, including, for example, contributions from countries such as India, Thailand or some countries from the Middle East. They seemed to have a similar starting point as the

“Coloniality”-type narrative patterns, but a much more “comfortable” attitude towards digital technologies.

These narrative patterns are characterised by seeing ecology (or the community/society) as the primary reference point and themselves as a (originally rather passive) beneficiary of digital technologies, and by having few ethical concerns; any such concerns are likely to be focused on prosperity and particular other regions or states. The emphasis is put on organic evolution, and on culture or education as the primary tools of governance.

Fundamental ethical assumptions	harmony/virtue	deontological	
	role-adequacy	material equality	
The protagonist and its role vis-à-vis technology	ecology	community/society	
	beneficiary/convenience		
The conflict and the antagonist	prosperity/marginalisation		
	government(s)		
The wider plot (what should happen in terms of governance)	culture/education		
	organic evolution		

Figure 5: “Beneficiary” pattern

However, we have also seen in these narrative patterns a recent tendency of actors to see themselves more as “active agents” and to place more trust in technology and the

economy as tools of governance, so “Beneficiary”-type narrative patterns may be disposed to evolve into “Actor”-type narrative patterns.

Fundamental ethical assumptions	harmony/virtue	deontological	
	role-adequacy	material equality	
The protagonist and its role vis-à-vis technology	ecology	community/society	
	beneficiary/convenience		actor/opportunities
The conflict and the antagonist	prosperity/marginalisation		
	government(s)		
The wider plot (what should happen in terms of governance)	culture/education		technology/economy
	organic evolution		self-determination

Figure 6: “Beneficiary” pattern turning into an “Actor” pattern

“HARMONY-OPPORTUNITY”-TYPE NARRATIVE PATTERNS

A number of different narrative patterns, reflected particularly in contributions from participants associated with the East Asian region, share as a starting point that the “good” is determined by some notion of “harmony”, that they tend to see “role adequacy” as the starting point for “fairness” and that their point of reference tends to be “ecology” (understood as the animate and inanimate world). What is possibly most striking about this narrative pattern, that marks it

out, in particular, from the “Coloniality” narrative pattern described above, is that this starting point is combined with a strong sense of being an actor with active agency in the digital transformation, seizing opportunities for a better future. That said, details vary considerably. Some contributors associated with Japan, for example, felt that concerns (if present at all) would mainly arise from the actions of malicious users. More generally characteristic was the emphasis on organic evolution, with a lot of trust placed in governments, and with the most appropriate governance tool being seen as (corporate) culture and education.

Fundamental ethical assumptions	harmony/virtue		
	role-adequacy		
The protagonist and its role vis-à-vis technology	ecology		actor/opportunities
The conflict and the antagonist		safety/harm	
			(other) users
The wider plot (what should happen in terms of governance)	culture/education		
	organic evolution		

Figure 7: Japan-associated variant of the “Harmony-Opportunity”-type pattern

A slightly different narrative pattern can be seen in contributions from participants associated with China, where emphasis is placed on prosperity and growth, and where the strategic focus is clearly on the “determination of the

able”. This context can thus be seen as orienting towards an approach in which the government makes decisions in the (perceived) general interest.

Fundamental ethical assumptions	harmony/virtue	
	role-adequacy	
The protagonist and its role vis-à-vis technology	ecology	
		actor/opportunities
The conflict and the antagonist	prosperity/marginalisation	
	government(s)	
The wider plot (what should happen in terms of governance)	culture/education	
		determination by the able

Figure 8: China-associated variant of the “Harmony-Opportunity”-type pattern

However, even in the context of this variant of the “Harmony-Opportunity” narrative pattern, a move might be identified towards seeing “ecology” as the victim and law and regulation as the appropriate governance tool, in

that big tech companies are increasingly seen as a threat to traditional (in this case “Chinese”) values. This underlines a more general trend we have seen in the AGIDE project with regard to the regulation of digital technologies.

Fundamental ethical assumptions	harmony/virtue	
	role-adequacy	
The protagonist and its role vis-à-vis technology	ecology	
		victim/risk ← actor/opportunities
The conflict and the antagonist	prosperity/marginalisation →	safety/harm
	government(s) →	tech industry
The wider plot (what should happen in terms of governance)	culture/education →	law/regulation
		determination by the able

Figure 9: Recent developments with regard to the China-associated variant of the “Harmony-Opportunity” narrative pattern

“SILICON VALLEY”-TYPE NARRATIVE PATTERNS

A distinctive approach was reflected in some contributions that could be characterised as a “Silicon Valley”-type

approach, with a very distinct narrative pattern. This narrative pattern can be characterised as clearly consequentialist, relying on formal equality, and super-individualist. It sees the individual as an actor, with a strong focus on

autonomy, concerns (if any) about malicious users of technology, self-determination as a method of change, and a lot of trust placed in big companies, technology and the economy as the major basis for legitimacy. This particular

narrative pattern might be seen to have provided an ideal ground for digital innovation within a capitalist accumulation-oriented economic system.

Fundamental ethical assumptions		consequentialist
		formal equality
The protagonist and its role vis-à-vis technology		individual(s)
		actor/opportunities
The conflict and the antagonist		autonomy/lack of freedom
		(other) users
The wider plot (what should happen in terms of governance)		technology/economy
		self-determination

Figure 10: Original “Silicon Valley”-type approach

At the same time, we were witness to signs that voices warning of risks to the individual and calling for more regulation are gaining ground. Yet, it is questionable whether this shift is strong enough to change fundamental mainstream beliefs. The following depicts what such

a movement might look like, particularly among expert groups: shifting towards a more critical attitude towards the technology industry and a call for governments to take control through regulation.

Fundamental ethical assumptions		consequentialist
		formal equality
The protagonist and its role vis-à-vis technology		individual(s)
	victim/risk	actor/opportunities
The conflict and the antagonist		autonomy/lack of freedom
	tech industry	(other) users
The wider plot (what should happen in terms of governance)		technology/economy
	law/regulation	self-determination

Figure 11: Moving away from the traditional “Silicon Valley” pattern?

“Silicon Valley-type narrative patterns have been quite influential, and we find them in different parts of the world. Sometimes they are modified by external factors.

For example, contributions from Israel seemed to suggest that the starting point would normally have been a traditional “Silicon Valley” narrative pattern. However,

we observed a narrative pattern that differs significantly from the original narrative pattern, in that the community seems to have overtaken the individual as the primary point of reference (“protagonist”). In a context of ongoing regional conflict, this community is seen as an actor actively engaging with emerging digital technologies, embracing new opportunities, particularly for the sake

of security, and fearful of the possible role of malicious actors. In terms of decision-making, the sense of living under threat has led to a paternalistic attitude that tolerates a substantial degree of surveillance, while trusting the government and the intelligence services as actors and choosing technology and the economy as the preferred tools of governance.

Fundamental ethical assumptions		consequentialist
		formal equality
The protagonist and its role vis-à-vis technology	community/society	←
		actor/opportunities
The conflict and the antagonist	safety/harm	←
		(other) users
The wider plot (what should happen in terms of governance)		technology/economy
	determination by the able	←

Figure 12: A typical (more secular) Israel-associated pattern

“GDPR”-TYPE NARRATIVE PATTERNS

Another characteristic narrative pattern seemed to emerge from a number of contributions, many of which come either from EU countries or from individuals expressing admiration for the EU’s digital policies (see also Bradford 2012). For lack of a better term, this narrative pattern is referred to as the “GDPR”-type narrative pattern. Contributions exhibiting this pattern took a distinctly deontological starting point (repeatedly mentioning fundamental rights and values as the foundation and benchmark for any governance decision), with material equality as a clear fairness goal and the individual as the primary reference point.

It is characteristic of these contributions that the individual is seen as primarily at risk from emerging digital technologies, with a focus on risks to autonomy (such as the right to informational self-determination or data protection) or, in fact, to democracy – seen as the very basis for citizens’ rights. It is likewise characteristic of this type of narrative that the risks are perceived as largely emanating from technology companies and their ever-growing power. Law and regulation (to be enacted by the State or multi-national institutions) are the clearly preferred means of governance. What gives the solutions provided by these laws and regulations legitimacy (apart from the democratic process that leads to them) is mainly individual self-determination (e.g. in the form of informed consent or “digital sovereignty”).

Fundamental ethical assumptions		deontological
		material equality
The protagonist and its role vis-à-vis technology		individual(s)
	victim/risk	

The conflict and the antagonist		autonomy/lack of freedom
		tech industry
The wider plot (what should happen in terms of governance)		law/regulation
		self-determination

Figure 13: The traditional “GDPR”-type pattern

CHARACTERISTIC NARRATIVE PATTERNS – MOVING TO THE MICROPERSPECTIVE

As the matrix of ethical narratives on digital ethics was necessarily over-generalising, we now move to a more meso-micro level, benefiting from the insights and examples of AGIDE working group members from relevant regions. Thus, the following examples are based on long-standing research by AGIDE working group members Christiane Wendehorst and Astrid Mager for the example of the European Union, as well as working group members Payal Arora and Manisha Pathak-Shelat for the example of India.

The data presented in the micro-perspective show that, perhaps contrary to what one might expect, the EU has very different and also divergent approaches, issues and concerns when it comes to technology and individual member states. Secondly, the example of India illustrates that there are differing narratives around surveillance by tech companies, surveillance by the government in different contexts, and surveillance by family, peers and employees, especially between the older and younger generations.

DIFFERENCES WITHIN A REGION – THE EXAMPLE OF THE EUROPEAN UNION

In the EU, the “GDPR”-type narrative pattern has been the dominant approach for decades. However, while this approach has been widely shared, if only unconsciously, by the populations in many EU member states, closer examination reveals that it has never been fully representative of EU countries or Europe as a whole (see also Bradford, 2012 & 2023; Pagallo, 2024).

Variations among EU Member States

So-called “Nordic” countries in the EU, for example, seem to be characterised by a clearly different, more consequentialist and, at the same time, more community-focused approach. This approach tends to emphasise prevention of harm rather than autonomy, to be much less critical of the technology industry and to be generally hostile towards regulation, preferring “soft” instruments (such as ombudspersons) and education as governance tools (cf., e.g., Kaun, Larsson and Masso, 2023, 327; Stråth and Wodak, 2009).

Fundamental ethical assumptions		consequentialist
		material equality
The protagonist and its role vis-à-vis technology		community/society
		actor/opportunities
The conflict and the antagonist		safety/harm
		(other) users
The wider plot (what should happen in terms of governance)	culture/education	
		determination by the able

Figure 14: Traditional approach in the “Nordic countries”

But things may be changing. Denmark, for example, has long been at the forefront of this “Nordic” pattern. However, several scandals, in particular the “Chromebook Scandal” (where it turned out that schoolchildren’s personal data had been handed over to Google and its parent company,

Alphabet) and the emergence of concerns relating to scoring of parents and facial recognition in public spaces, have had an impact on such narrative patterns in Denmark. A related shift in public opinion has brought Denmark more or less in line with the “GDPR”-type pattern.

Fundamental ethical assumptions			consequentialist
		material equality	
The protagonist and its role vis-à-vis technology		community/society	individual(s)
		victim/risk	actor/opportunities
The conflict and the antagonist		safety/harm	
		tech industry	(other) users
The wider plot (what should happen in terms of governance)	culture/education	law/regulation	
		determination by the able	self-determination

Figure 15: Denmark after the “Chromebook” and other scandals

Historically, trust in governmental structures has been very low in Central and Eastern European (CEE) countries (e.g. Jansen, 2023, 4–5; Centraal Bureau voor de Statistiek, 2023; Horakova, 2020). Accordingly, the focus of concern

associated with these countries is more on malicious public and private actors abusing technology for surveillance purposes, with more trust being put in big companies and technology than in government and regulation.

Fundamental ethical assumptions		deontological	
		material equality	
The protagonist and its role vis-à-vis technology			individual(s)
		victim/risk	
The conflict and the antagonist			autonomy/lack of freedom
	government(s)		
The wider plot (what should happen in terms of governance)			technology/economy
			self-determination

Figure 16: A government-sceptic pattern from CEE countries

The United Kingdom after Brexit

It is interesting to see that the narrative patterns of ethical narratives in the UK, as a country where the GDPR has

been in force for a number of years, have recently shifted towards a “Silicon Valley” pattern.

Fundamental ethical assumptions			consequentialist
		material equality	→ ? formal equality
The protagonist and its role vis-à-vis technology			individual(s)
		victim/risk	→ ? actor/opportunities
The conflict and the antagonist			autonomy/lack of freedom
		tech industry	→ ? (other) users
The wider plot (what should happen in terms of governance)		law/regulation	→ ? technology/economy
			self-determination

Figure 17: Post-Brexit UK – moving from the “GDPR” to the “Silicon Valley” pattern?

However, it is still unclear how far developments will go in this direction, i.e. whether dominant narratives in the UK will remain close to the “GDPR”-type pattern with just some twists and modifications, or whether the dominant narrative will actually change entirely.

Room for Alternative EU Narratives?

While the EU itself (i.e. its institutions and, in particular, the EU legislator) has defended and buttressed the “GDPR”-type pattern for a long time, and while this can still be seen as the hegemonic narrative, things seem to be changing. With the Artificial Intelligence Act (AIA), the legislator has moved away from a focus on autonomy and towards

a risk-based approach that focuses primarily on safety from harm. The “risk-based approach” (cf. European Commission, 2023; Mahler, 2021) of the AIA seems to be challenged by many, but to be by and large accepted. In particular, this seems to fit in well with the emphasis on “victim/risk”. It does not come as a surprise, though, that the absence of individual rights in the AIA proposal was criticised (Wendehorst, 2021; Prainsack and Forgó, 2024; Mladenov, 2023) and that the only individual right added by the European Parliament during the legislative process related to the explainability of AI – i.e. a right closely related with autonomy.

Fundamental ethical assumptions		deontological	
		material equality	
The protagonist and its role vis-à-vis technology			individual(s)
		victim/risk	
The conflict and the antagonist		safety/harm	← ? autonomy/lack of freedom
		tech industry	
The wider plot (what should happen in terms of governance)		law/regulation	
			self-determination

Figure 18: Risk-based approach of the AI Act – few changes required

In recent legislation on the data economy (such as the Data Governance Act and the Data Act; see: Cohen and Wendehorst, 2023), the EU legislator has tried to move towards a more community-oriented approach by liberating data for the European data economy, thereby seeming to turn some of the basic assumptions of the GDPR (such as data minimisation or purpose limitation) on their head (e.g. Schütte, 2024; Sartor and Lagioia, 2020). However, it proved to be unacceptable to openly challenge, for example, the focus on individual rights or the focus on self-determination, as the “GDPR”-type narrative is still too strong. This has led the EU legislator to present legislative measures in the field of data economy law as “user

empowerment” (Wendehorst, 2023) or similar, putting all the burden on the individual user and favouring, e.g., individual data portability rights over targeted data access rights and data intermediaries and data altruism organisations over supervised data spaces. It is only in the proposal for a European Health Data Space that self-determination seems to be replaced to a certain extent by a more community-oriented approach. This goes into the direction of a “data solidarity” narrative (a term coined by Prainsack *et al.*, 2022) – with unclear prospects, as the proposal is under substantial attack and some Member States are fighting fiercely for requirements for informed consent or at least an opt-out mechanism.

Fundamental ethical assumptions		deontological	
		material equality	
The protagonist and its role vis-à-vis technology		community/society	← X individual(s)
		victim/risk	X → actor/opportunities
The conflict and the antagonist	prosperity/marginalisation	← X → autonomy/lack of freedom	
		tech industry	
The wider plot (what should happen in terms of governance)		law/regulation	
		determination by the able	← X self-determination

Figure 19: Data economy legislation – change of the law but not of the narrative

Some potential EU digital ethics narratives would be plausible, but remain rather marginalised (Mager, 2023). These include the “EU digital sovereignty” narrative, which could ideally be the European counterpart to the “Coloniality”-type pattern (see above). It emphasises the need for the EU’s economic prosperity and technological advancement to be as independent as possible from other regions of the world and to avoid becoming a “digital colony” of big tech nations (Lewandowski, 2014; Baur, 2023; Mager, 2023). Various attempts have been made to establish this narrative – and start corresponding action, such as by developing a European cloud ecosystem called GAIA-X, for example. The European flagship project GAIA-X is framed as “Europe’s moon shot”, but also in terms of a geopolitical fight for “European sovereignty” in the IT sector (Baur, 2023). The German Ministry for Economic Affairs and Climate Action describes GAIA-X as follows: “The goal is a secure and federated data infrastructure that stands for European values, digital sovereignty of the data owners, interoperability of different

platform[s] and open source.” (compare: homepage of BMWK, 2024).⁴ This description serves to show that even where digital sovereignty should be the main issue, the focus of the narratives remains on individuals and their autonomy, reinforcing the dominance of the “GDPR” narrative pattern and foregoing opportunities to establish truly alternative patterns.

Similar notions of European sovereignty are mobilised in current search engine projects. The EU-funded project “European Web Search” formulates its central goal as “promoting Europe’s independence in Web Search” according to its website. The joint initiative of 14 European research and computing centres aims at building an open, independent web index so that different search engines and applications can be built on top of it (Lewandowski, 2014 & 2019; Mager, 2014 & 2023). It thereby aims at contributing to “Europe’s digital sovereignty as well as promoting an open human-centered search engine market”, as can be further

⁴ <https://www.bmwk.de/Redaktion/EN/Dossier/gaia-x.html>

read on the website. An open web index would also help search engines that currently partner with big, corporate search engines like Bing (for using their index and results)

to become independent from big tech companies. Whether such projects – and the corresponding narrative of digital sovereignty – will grow in the future remains to be seen.

Fundamental ethical assumptions		deontological	
		material equality	
The protagonist and its role vis-à-vis technology		community/society	individual(s)
		victim/risk	
The conflict and the antagonist	prosperity/marginalisation		autonomy/lack of freedom
	government(s)	tech industry	
The wider plot (what should happen in terms of governance)		law/regulation	technology/economy
			self-determination

Figure 20: Hypothetical “EU digital sovereignty” narrative

DIFFERENCES WITHIN A POPULATION – THE EXAMPLE OF INDIA

Patterns of narratives also vary across a population. AGIDE data provide insights into the distribution of narrative patterns, for example, for India, but similar findings can be formulated for many other countries. Variations in narrative patterns in India can be observed based on generational experience and aspirations among other factors, and for every aspect of digital ethics, such as surveillance, and the role of technology companies (Arora, 2016 & 2019a; Komarraju, Arora and Raman, 2022). There are huge differences and debates around surveillance and also different meanings

attached to the term based on one’s “situatedness”. The same individuals also look differently at ethical aspects of surveillance by tech companies, surveillance by the government in different contexts, and surveillance by family, peers, and employees (Bhatia, Arora and Pathak-Shelat, 2021). For example, upwardly mobile youth may focus on benefits, individual autonomy, and faith in tech companies with education as the tool of governance (Bhatia and Pathak-Shelat, 2023; Bhatia, Pathak-Shelat and Arora, 2024). At the same time, they want the government to play a protective role with respect to individual privacy, safety, and security. Data privacy and data protection are important concerns for them (cf. Arora, 2019b).

Fundamental ethical assumptions			consequentialist
		material equality	
The protagonist and its role vis-à-vis technology			individual(s)
	beneficiary/convenience	victim/risk	
The conflict and the antagonist			autonomy/lack of freedom
	government(s)	tech industry	
The wider plot (what should happen in terms of governance)	culture/education	law/regulation	
			self-determination

Figure 21: India – upwardly mobile youth

The attitudes and concerns of some older adults in India may show a pattern focusing on the need for protection from harm and self-perception as “victims”. They may consider community as a primary reference point with suspicion towards tech companies for causing cultural

disruption. Such adults value regulation as the tool of governance. In addition to self-determination, the role of the state in ensuring access to and affordability of technology is also important as part of the overall concept.

Fundamental ethical assumptions			consequentialist
	role-adequacy		
The protagonist and its role vis-à-vis technology		community/society	
		victim/risk	
The conflict and the antagonist		safety/harm	
		tech industry	
The wider plot (what should happen in terms of governance)		law/regulation	
		determination by the able	

Figure 22: India – older generation

The Indian government itself heavily pushes a discourse with the Indian nation at the centre, individuals as beneficiaries and actors, and a mixed approach to tech companies and education as well as regulation as the tool of

governance. The government also demonstrates a strong aversion to those seen as posing a risk to national security, the dominant culture and the safety of young people.

Fundamental ethical assumptions			consequentialist
		material equality	
The protagonist and its role vis-à-vis technology		community/society	
			actor/opportunities
The conflict and the antagonist		safety/harm	
		tech industry	(other) users
The wider plot (what should happen in terms of governance)		law/regulation	
		determination by the able	

Figure 23: India – official government discourse

There are also some think tanks, groups of intellectuals and civil society organisations that challenge these narrative patterns by foregrounding the importance of equity, community and autonomy, showing a strong resistance to

the power of tech companies and pushing for the government’s role in creating material equity and regulations to control harm (cf. Pathak-Shelat and Bhatia, 2019).

Fundamental ethical assumptions	harmony/virtue		
		material equality	
The protagonist and its role vis-à-vis technology		community/society	
		victim/risk	
The conflict and the antagonist			autonomy/lack of freedom
		tech industry	
The wider plot (what should happen in terms of governance)		law/regulation	
		determination by the able	

Figure 24: India – public interest groups and intellectuals

PART D: IMPLICATIONS AND OUTLOOK

The AGIDE project provides unique insights into how digital ethics are discussed in different parts of the world and in various contexts. Taking a “situated” approach allowed the project to focus on local knowledge contexts and understand the complexity and diversity of narratives by looking at their positionality, as discussed in Part A (Haraway, 1988). The project team set out to learn from experts about the values, principles and concerns shaping discourses on digital ethics in different regional, national and institutional contexts. As noted, insights in this report draw upon two main sources: interviews with 75 individuals in a total of 28 countries and three workshops, where experts from different world regions presented their perspectives, answered questions that had arisen from the interview data and reviewed and discussed preliminary insights of the project (see Part B). AGIDE cannot, and does not, claim to represent all relevant positions (see also Part B for more details on the methodology and limitations), and its focus has been on professional experts rather than people’s perspectives as citizens. Nevertheless, the AGIDE approach has led to several findings that seem sufficiently robust to treat them as initial insights for further exploration. In this section, we will summarise these insights and suggest what we consider fruitful ways forward.

THE POWER OF ETHICAL NARRATIVES

- AGIDE set out with the observation that the relative ease with which countries around the world seem to agree on universal principles of digital ethics along the lines of “fairness”, “transparency” and “accountability” seems to contrast sharply with the vast differences in technology adoption that we see around the world, and also the vast differences in attitudes towards technology. **The AGIDE project found that there is a remarkable consistency in core values (such as “justice”, “dignity” or “privacy”) across different regions of the world.** Contrary to initial expectations that the attitudinal and regulatory differences explored by AGIDE might result from discernible differences in emphasis on particular values, the data collected did not support such distinctions.
- However, **AGIDE’s research showed that major differences lie in something else: in the narratives of digital ethics.** Narratives are stories that are told repeatedly, consisting of a series of events that are selected and arranged in a particular order, often including central characters (protagonists, antagonists), a conflict and a plot. The **“matrix of digital ethics narratives” approach**, from which we in turn derived the *ethical narrative patterns* detailed above, attempts to provide a framework within which views and concerns can be meaningfully articulated in relation to eight key dimensions. Taken together, the very particular views and approaches to these eight issues result in characteristic patterns of storytelling that reflect certain characteristic digital ethics narratives. From a macro-perspective, the characteristic pattern types (or archetypes) that emerged include the “Coloniality”-type pattern, the “Beneficiary”-type pattern, the “Harmony-Opportunity”-type pattern, the “Silicon Valley”-type pattern, and the “GDPR”-type pattern. Exploring the narratives from a micro-perspective, looking at specific regions or populations, revealed a much greater diversity– highlighting both the nuanced nature of narratives around digital ethics in context, but also showcasing the analytical sharpness and possibility for comparison purchased by using the common framework of the matrix.
- While the core values underpinning digital ethics are very similar across world regions, **there are considerable differences between world regions (and sometimes also countries) in the ontological positions, heuristics, and historical experiences, as well as legal and religious traditions, that give shape to digital ethics debates** and narratives (see Part C). One example is the deontological approach in the EU, with its strong focus on the individual, whose rights must be defended against the tech industry. Against this backdrop, the EU’s focus has been on ensuring digital self-determination, with binding supranational regulation as the preferred governance tool. This puts the EU in marked contrast to the US, for example, which also focuses on the individual and their right to self-determination, but

more strongly from a consequentialist perspective. The US approach emphasises entrepreneurial opportunity and places trust in big companies, technology and the market. Narratives in other countries such as Egypt, for example, focus on safety and protection of citizens from specific digital technologies, such as surveillance cameras.

- **Narratives of digital ethics do not correspond simply with world regions or “cultures”.** *A-priori* assumptions about supposed “cultural differences” cannot be confirmed by our data. Moreover, they risk rendering invisible the great heterogeneity within countries and so-called or associated “cultures”. An overly simplistic view of ethical approaches mapping onto “cultures” also draws attention away from the ethical concerns that emerge from the inequitable distribution of harms and benefits. In addition, **concerns and solutions discussed in digital ethics differ not only between regions (and, sometimes, countries), but also between sectors and population groups**, and can thus be individual- or group-specific, context-specific, and technology-specific. For example, privacy or justice – to name just two examples – are important concerns across the globe. Yet, the relative priority given to them and the governance tools that are considered adequate to protect them are different for specific sectors (e.g. health, environmental protection, work) and often also for specific groups (e.g. according to gender, age, socio-economic status).
- **The primary point of reference (the “protagonist” of the story) proved to have significant influence on the rest of the narrative.** To a certain extent, this may be seen to confirm the existing stereotype about different emphasis put on the individual (as a stereotypical “Western” or “Global North” perspective) or on the community (as a stereotypical “Eastern” or “Global South” perspective), respectively, and the same could be said about the overlapping stereotype about rights-based and duty-based approaches. However, **findings were much more nuanced than these stereotypes or tropes.** Not only are there significant differences between narratives that take “ecology” – understood as humans and both the animate and the inanimate environment – as a starting point and narratives that focus exclusively on human communities or societies. There are also immense differences as to the role the “protagonist” is seen to play relative to technology, the nature of the main concern with regard to this role, and the identity of the perceived main “antagonist”. A focus on “ecology” or “community” can therefore mean very different things, depending on whether, for example, the community is perceived as being marginalised by powerful actors in other world regions, or whether it is perceived as actively embracing technology to fight against harmful activities threatening the community.
- It was particularly interesting to observe that many **narrative patterns are fluid to some extent, and that there are remarkable shifts as far as dominant narrative patterns in a country or region are concerned.** According to expert contributors to the AGIDE project, for example, it is possible to observe major shifts in narrative on the Asian continent, where we see “Beneficiary”-type narrative patterns develop into “Harmony-Opportunity”-type narrative patterns, but also a new China-associated narrative about the technology industry in the “antagonist” role, whose medium- and long-term effects on China’s position remain to be seen.
- However, **some narratives showed persistence, so deeply entrenched that even forceful government initiatives alone may not be enough to bring about change.** The EU was presented above as an illustrative example where the “GDPR”-type narrative pattern resists alteration to an extent that might potentially prevent EU policymakers from effectively bringing about change. It is difficult to say why exactly some narrative patterns are so strongly embedded. Coming back to the example of the EU, one of the reasons could be that, in the narrative, a number of elements of the plot elsewhere in the matrix (an emphasis on the individual, on autonomy, and on self-determination) have been inextricably linked with the basic starting point of a deontological conception (or “notion”) of the “good”, which, in the EU, means a firm rooting in fundamental rights. This alone, however, does not explain why placing the technology industry always in the “antagonist” role, is such a persisting element.
- **Situated debates and narratives about digital ethics have a profound impact on the way digital technologies are perceived and dealt with.** This is true in particular for dominant narratives that are enforced by strong political forces or deeply rooted in fundamental beliefs and underpinned by societal taboos. These narratives affect where and how resources are spent, what policy makers see as problems to be solved, and where human dignity and freedoms are seen to be challenged. If there is a global trend, it is one towards law and regulation as a governance tool – but based on very different narratives and therefore different strategies (Bradford, 2023; Pagallo 2024). Characteristic differences in solutions exist, with anti-industry sentiments increasingly driving government intervention and regulatory measures.

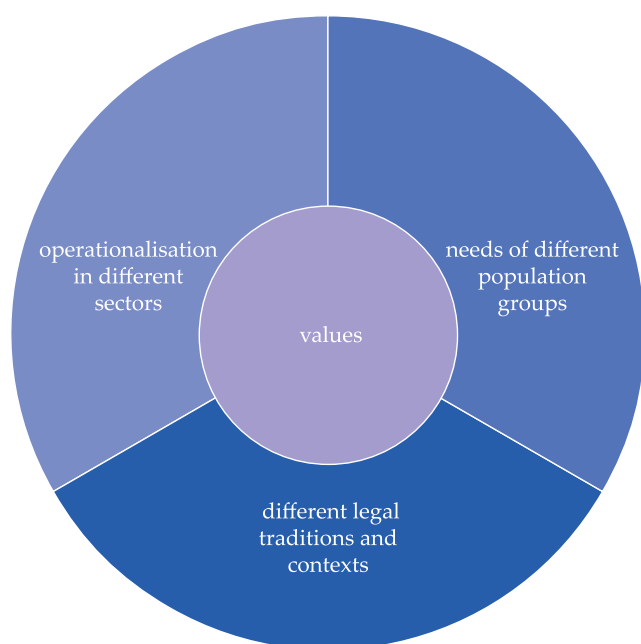


Figure 25: Spheres of factors influencing digital ethics narratives [source: Working Group Members]

IMPLICATIONS FOR FURTHER RESEARCH AND POLICY-MAKING

- Together, the findings and analyses of Part B and Part C of this report highlight the need for further research to explore whether the narratives are **causes or symptoms of the differences** we perceive, or both. Further research is also needed into the factors that contribute to the development of specific narratives, both at the macro- and meso-/micro-level, and we also need to understand better **the conditions that are conducive to the transformation of established narratives or that cause established narratives to resist even major shifts in the policies** pursued, potentially hindering important policy changes. The influence of narratives in “digital spaces” on political discourse and governments themselves could also usefully be the subject of further research (Budka, 2019; Page, 2018; Palmberger and Budka, 2020).
- In addition, several steps could be taken to **promote engagement with digital ethics narratives**: in terms of **academic research**, firstly, **longitudinal analyses of narratives** to track their evolution in the context of digital transformation would be needed to gain a better understanding of their impact on technology development, adoption, and regulation. Secondly, **an impact analysis of narratives** on stakeholder engagement, such as on policymakers, engineers, and researchers, would allow us to gain more insights into the power structures involved in technology development, which is needed to foster **responsible innovation** (cf., e.g., Reijers and Coeckelbergh 2020, 200–202). Thirdly, in a similar vein, more empirical evidence is needed on the impact of narratives on individual technology usage and lived experiences within particular socio-cultural contexts. Fourth, research on how the results of narrative analysis can be integrated into the actual development of **ethical AI systems and algorithmic decision-making processes** is essential in order to achieve responsible and ethical AI.
- Narrative analysis also needs to be integrated into **policy formulation and technology development**, as it would increase awareness of established (dominant) narratives that implicitly steer policy discourses. Considering the power dynamics of narratives and **strengthening inclusive narratives** would also be the next step to more equity and fairness. In this process, it is important to involve different stakeholders, including communities, businesses, and academia, in the **narrative-shaping process** wherever and whenever possible and appropriate, from local to national to international levels. Similarly, policies should be regularly reviewed and adapted based on narrative dynamics. Finally, it is of utmost relevance to invest in **digital literacy and narrative education** so that all affected stakeholders and, above all, representatives in decision-making positions, understand how they shape their perspectives on technologies and digital ethics and so that they can potentially co-create alternative narratives.
- Reflecting on (dominant) digital ethics narratives as drivers and tools for establishing values and principles, and on the power dynamics involved, means acknowledging that **digital ethics is a deeply political practice**, both within countries and across the globe. It is therefore necessary not only to consider one’s role in the co-creation of narratives and the negotiation of power, but also to take digital ethics seriously when it comes to the political implications of ethical assessments and demands. AGIDE agrees with the call for a “re-politisation” (van Maanen, 2022) and a “political turn” of digital ethics (Coeckelbergh, 2022), which requires ensuring that **ethical principles can be translated meaningfully into practice, backed up where necessary with policies and legislation**.
- In addition to the points mentioned, shared values that are seen as morally, politically or even legally binding can help to protect people and ensure that minorities or oppressed groups are not harmed by norms benefitting the dominant or majority group. Thus, there was wide agreement among AGIDE participants and contributors that such **shared values can – and should – prescribe (at least a minimum) standard for digital ethics. At the same time, it is important to note that situated values**

and principles of digital ethics must be given equal consideration. This means that research and policy should attend to ways in which cultural values, social and economic aspects, political regimes, centre/periphery dimensions and so on, play into digital ethics. However, participants warned us against idealising respect for particularistic values under the guise of “cultural difference” as this could be used as an excuse to uphold harmful values.

- Additionally, unequal access to resources is one of the major factors shaping digital and other divides in and across societies. Thus, **an important objective for digital ethics is firstly to work towards more equitable (material) conditions of technology development and innovation, and secondly, a more equitable distribution of both benefits and harms** – within societies, across societies and between private businesses and citizens. Rather than more principle-based guidelines for digital ethics, what is needed are analyses of what the realisation of values and principles entails in specific practical contexts, for specific technologies and regarding the needs of specific population groups.
- Therefore, what we may need for a good digital future are **technologies that are genuinely developed in as many parts of the world as possible, and from as many perspectives as possible, provided that those technologies remain true to the shared values mentioned above.** In order to achieve this, communities in all parts of the world must be empowered and enabled to develop their own digital technologies. This requires such basic things as access to computing power, electricity, and appropriate data. However, it also requires digital literacy and education that is sufficient for actively shaping technologies according to the views and needs of the relevant communities.
- Finally, **it is crucial to understand how diversity can help to increase resilience.** We need to consider the often-overlooked potential of the diversity of knowledge that has been shaped in the political and economic spheres in different geopolitical contexts, and which also reflects differences in the education systems that educate citizens as members of particular nations. To date, the focus has mainly been on initiatives to increase

diversity in different parts of the digital sector, including in the design and development workforce, in management, and in oversight mechanisms. While this can certainly help, the diverse inputs are no longer visible in the resulting technology, which tends to be **an amalgam of a few dominant voices** and interests. To address these issues, **plurality - or “pluriversality” - may be the more important concept in the long run**, because it ensures that different voices remain distinct.

CLOSING REMARKS

The AGIDE project was a unique opportunity to explore digital ethics from the perspective of experts across the globe. While AGIDE cannot, and does not, claim to be representative of the views of “the people” in all these world regions, it makes a contribution to mapping and analysing digital ethics from a pluriversal perspective. The findings of the AGIDE project, which were obtained inductively from 75 interviews with experts, as well as workshops and the analysis of relevant literature, paint a rather nuanced picture: While the concerns and values that are articulated within digital ethics discourses do not map neatly onto different world regions, religions, or “cultures”, there are considerable differences between regions and countries regarding the ontological positions, as well as relevant traditions, that shape digital ethics narratives. Additional levels of variation seem to exist at the level of different population groups, such as younger and older people, different genders, and different socio-economic strata. These group-related differences can cut across national and regional borders.

While the AGIDE project has made a start in mapping and analysing digital ethics debates at the global level, it can only be a start. As this final section of the report in particular suggests, further dialogue and research would be needed to explore in more empirical detail similarities and commonalities within and across countries. Moreover, especially against the backdrop of recent developments such as the spread of generative AI, debates are necessary at all levels – regional, national, and international – about what instruments and approaches are needed to protect the fundamental rights and basic needs of all people. While these rights and needs are the same as in the paper age, in the digital era, new tools are needed for this purpose.

REFERENCES

- Adam, B. and Groves, C. (eds.) (2007). *Future matters: action, knowledge, ethics*. Leiden: Brill (Supplements to the study of time, 3).
- Aggarwal, N. and Floridi, L. (2020). 'Introduction to the Special Issue on Intercultural Digital Ethics', *Philosophy and Technology*, 33(4), pp. 547–550.
- Aho, B. and Duffield, R. (2020). 'Beyond surveillance capitalism: Privacy, regulation and big data in Europe and China', *Economy and Society*, 49(2), p. 187–212.
- Anderson, B.R.O. (1983). *Imagined communities: reflections on the origin and spread of nationalism*. Rev edn. London: Verso.
- Arendt, H. (1958). *The Human Condition*. Reprint, Chicago: University of Chicago Press, 2013.
- Arora, P. (2016). 'Bottom of the data pyramid: Big data and the global south', *International Journal of Communication*, 10, pp. 1681–1699.
- Arora, P. (2019a). *The next billion users: Digital life beyond the West*. Cambridge: Harvard University Press.
- Arora, P. (2019b). 'Decolonizing privacy studies', *Television & New Media*, 20(4), pp. 366–378.
- Austrian Advisory Board (2024). *Fachbeirat Ethik der Künstlichen Intelligenz. Expertise für die Umsetzung der UNESCO-Empfehlung*. Available at: <https://www.unesco.at/wissenschaft/wissenschafts-und-bioethik/ethik-der-kuenstlichen-intelligenz/fachbeirat-ethik-der-kuenstlichen-intelligenz>. (Accessed: 12 April 2024).
- Awad, E. et al. (2018). 'The Moral Machine Experiment', *Nature*, 563(7729), pp. 59–64.
- Bamberg, M. (2012). 'Narrative Analysis', in Cooper, H. (ed.) *APA Handbook of Research Methods in Psychology*. Washington: American Psychological Association (Quantitative, Qualitative, Neuropsychological, and Biological, 2), p. 77–94.
- Bamberg, M. and Georgakopoulou, A. (2008). 'Small Stories as a New Perspective in Narrative and Identity Analysis', *Text & Talk—An Interdisciplinary Journal of Language Discourse Communication Studies*, 28(3), pp. 377–96.
- Bareis, J. and Katzenbach, C. (2022). 'Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics', *Science, Technology, & Human Values*, 47(5), pp. 855–881.
- Baur A. (2023). 'European Dreams of the Cloud: Imagining Innovation and Political Control', *Geopolitics*, pp. 1–15.
- Berberich, N., Nishida, T. and Suzuki, S. (2020). 'Harmonizing Artificial Intelligence for Social Good', *Philosophy & Technology*, 33(4), pp. 613–638.
- Bhatia, K.V., Arora, P. and Pathak-Shelat, M. (2021). 'Good girls don't go online: Unpacking the quotidian playful resilience influencing girls' social and digital engagements', *International Journal of Communication*, 15, pp. 4755–4773.
- Bhatia, K.V. and Pathak-Shelat, M. (2023). *Gen Z, Digital Media, and Transcultural Lives: At Home in the World*. Lanham: Lexington Books.
- Bhatia, K.V., Pathak-Shelat, M. and Arora, P. (2024). 'DIY education in the digital era: youth-driven learning strategies and curricula for the future of work opportunities', *Education and Information Technologies*, 29(2), pp. 1407–1426.
- Blok, V. and Lemmens, P. (2015). 'The Emerging Concept of Responsible Innovation. Three Reasons Why It is Questionable and Calls for a Radical Transformation of the Concept of Innovation', in Koops, B.-J. et al. (eds.) *Responsible Innovation 2: Concepts, Approaches, and Applications*. Dordrecht: Springer, pp. 19–35.
- BMWK (2024). *The Gaia-X Ecosystem – A Sovereign Data Infrastructure for Europe*. Available at: <https://www.bmwk.de/Redaktion/EN/Dossier/gaia-x.html>. (Accessed: 12 April 2024).
- Booth, C. et al. (2009). 'Scenarios and counterfactuals as modal narratives', *Futures: the journal of policy, planning and futures studies*, 41(2), pp. 87–95.
- Bogner, A., Littig, B. and Menz, W. (2009). 'Introduction: Expert Interviews – An Introduction to a New Methodological Debate', in Bogner, A., Littig, B. and Menz, W. (eds.) *Interviewing Experts*, page 281. London: Palgrave Macmillan, pp. 1–15.
- Braun, V. and Clarke, V. (2006). 'Using thematic analysis in psychology', *Qualitative research in psychology*, 3(2), pp. 77–101.
- Bradford, A. (2012). 'The Brussels Effect', *Northwestern University Law Review*, 107(1), pp. 1–68.
- Bradford, A. (2023). *Digital Empires: The Global Battle to Regulate Technology*. Oxford: Oxford University Press.

- Brown, N. and Michael, M. (2003). 'A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects', *Technology Analysis & Strategic Management*, 15(1), pp. 3–18.
- Budka, P. (2019). 'Von der Cyber-Anthropologie zur Digitalen Anthropologie. Über die Rolle der Kultur- und Sozialanthropologie im Verstehen soziotechnischer Lebenswelten', in Luger, M., Graf, F. and Budka, P. (eds.) *Ritualisierung – Mediatization – Performance*. Göttingen: Vienna University Press, pp. 163–188.
- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Bynum, T.W. (2015). 'Computer Ethics: Basic Concepts and Historical Overview', *Stanford Encyclopedia of Philosophy*, Winter 2005 Edition. Available at: <https://plato.stanford.edu/Archives/Win2005/entries/ethics-computer/>. (Accessed: 17 January 2024).
- Capurro, R. (2005). 'Privacy. An intercultural perspective', *Ethics and Information Technology*, 7(1), pp. 37–47.
- Capurro, R. (2008). 'Intercultural information ethics', in Himma, K. and Tavani, H. (eds.) *The Handbook of Information and Computer Ethics*. Malden: Wiley.
- Carroll, S.R. et al. (2022). 'Using Indigenous Standards to Implement the CARE Principles: Setting Expectations through Tribal Research Codes', *Frontiers in genetics*, 13, pp. 823309–823309.
- Cave, S., Kanta, D., and Dillon, S. (eds.) (2020). *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford University Press.
- Chaudhary, M.Y. (2020). 'Initial Considerations for Islamic Digital Ethics', *Philosophy & Technology*, 33, pp. 639–657.
- Cho, H., Rivera-Sánchez, M. and Lim, S.S. (2009) 'A multinational study on online privacy: global concerns and local responses', *New media & society*, 11(3), pp. 395–416.
- Coeckelbergh, M. (2019). *Moved by Machines: Performance Metaphors and Philosophy of Technology*. London: Routledge.
- Coeckelbergh, M. (2022). *The political philosophy of AI: an introduction*. Cambridge: Polity Press.
- Cohen, J. (2016). 'The Surveillance-Innovation Complex: The Irony of the Participatory Turn', in Barney, D. et al. (eds.) *The Participatory Condition in the Digital Age*. Minneapolis: University of Minnesota Press, pp. 207–226.
- Cohen N. and Wendehorst, C. (2023). *ALI-ELI Principles for a Data Economy: Data Transactions and Data Rights*. The American Law Institute and the European Law Institute (eds.) New York/Toronto: Thomson Reuters. https://europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ALI-ELI_Principles_for_a_Data_Economy.pdf.
- Collins, P.H. (2008). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.
- Couldry, N. and Mejias, U.A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press (Culture and Economic Life).
- Crenshaw, K. (1989). 'Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Policies', *University of Chicago Legal Forum*, (1989)1, pp. 139–167.
- Culler, J.D. (2011). *Literary theory: a very short introduction*. 2nd edn. Oxford: Oxford University Press.
- Centraal Bureau voor de Statistiek (2024). 'Trust in others, legal system and politics; European comparison'. Available at: <https://data.europa.eu/data/datasets/670-vertrouwen-in-mensen-recht-en-politiek-europese-vergelijking?locale=en>. (Accessed: 12 April 2024).
- Delaney, N. et al. (2020). *Science with and for Society in Horizon 2020. Achievements and Recommendations for Horizon Europe*. Brussels: European Commission. Available at: https://research-and-innovation.ec.europa.eu/document/download/a670a1ae-b084-4758-987e-fb9af495facc_en?filename=ec_rtd_swafs_achievements_and_recommendations_report.pdf. (Accessed: 14 January 2024).
- De Fina, A. and Georgakopoulou, A. (2008). 'Analysing narratives as practices', *Qualitative research: QR*, 8(3), pp. 379–387.
- Ess, C. (2006). 'Ethical pluralism and global information ethics', *Ethics and Information Technology*, 8(4), pp. 215–226.
- Ess, C. (2020). 'Interpretative Pros Hen Pluralism: from Computer-Mediated Colonization to a Pluralistic Intercultural Digital Ethics', *Philosophy & Technology*, 33(4), pp. 551–569.
- European Commission (2023). *Artificial Intelligence – Questions and Answers**. Available at: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683. (Accessed: 12 April 2024).
- Sartor, G. and Lagioia, F. (2020). *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. Brussels: European Union, EPRS | European Parliamentary Research Service, Scientific Foresight Unit (STOA), STUDY - Panel for the Future of Science and Technology.
- Firestone, W.A. (1993). 'Alternative Arguments for Generalizing from Data as Applied to Qualitative Research', *Educational Researcher*, 22(4), pp. 16–23.
- Fjeld et al. (2020). 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI', *Berkman Klein Center for Internet & Society*. Available at: <https://cyber.harvard.edu/publication/2020/principled-ai>. (Accessed: 16 January 2024).
- Fleck, L. (1981). *Genesis and Development of a Scientific Fact*. Translated by Bradley, F. and Trewn, T.J. Chicago: University of Chicago Press.
- Floridi, L. (2006). 'Information Ethics: Nature and Scope', *Acm Sigcas Computers and Society*, 36(2), pp. 21–36.
- Floridi, L. (2010). *The Cambridge Handbook of Information and Computer Ethics*. Cambridge: Cambridge University Press.

- Floridi, L. (2019). 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical', *Philosophy & Technology*, 32(2), pp. 185–193.
- Floridi, L. and Strait, A. (2020). 'Ethical Foresight Analysis: What it is and Why it is Needed?', *Minds and Machines*, 30(1), pp. 77–97.
- Floridi, L. and Taddeo, M. (2016.) 'What is data ethics?', *Philosophical transactions of the Royal Society of London. Series A: Mathematical, physical, and engineering sciences*, 374(2083), p. 20160360.
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Friedman, B. (ed.) (1997). *Human Values and the Design of Computer Technology*. Center for the Study of Language and Information. Chicago: Chicago University Press.
- Fuchs, C. (2022). *Digital Ethics: Media, Communication and Society*. Abingdon: Routledge.
- Golbin, I. and Axente, M.L. (2021). '9 ethical AI principles for organizations to follow', in *World Economic Forum*. Available at: <https://www.weforum.org/agenda/2021/06/ethical-principles-for-ai/>. (Accessed: 16 January 2024).
- Government of the UK (2020). *Data Ethics Framework*. London: Government Digital Service. Available at: https://assets.publishing.service.gov.uk/media/5f74a4958fa8f5188dad0e99/Data_Ethics_Framework_2020.pdf. (Accessed: 16 January 2024).
- Götsch, B. and Palmberger, M. (2022). 'The Nexus of Anthropology and Narrative: Ethnographic Encounters with Storytelling Practices', *Narrative culture*, 9(1), pp. 1–22.
- Groth, S. (2019). 'Political Narratives / Narrations of the Political: An Introduction', *Narrative Culture*, 6(1), pp. 1–18.
- Guay, R., & Birch, K. (2022). A comparative analysis of data governance: Socio-technical imaginaries of digital personal data in the USA and EU (2008–2016). *Big Data & Society*, 9(2). <https://journals.sagepub.com/doi/full/10.1177/20539517221112925>
- Halegoua, G. and Polson, E. (2021). 'Exploring 'digital place-making'', *Convergence (London, England)*, 27(3), pp. 573–578.
- Haraway, D. (1988). 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective', *Feminist Studies*, 14(3), pp. 575–599.
- Harding, S. (2006). *Science and Social Inequality: Feminist and Postcolonial Issues*. Champaign: University of Illinois Press.
- Hartsock, N.C.M. (2019). *The Feminist Standpoint Revisited, and Other Essays*. New York: Routledge.
- Heintz, M. (2009). *The Anthropology of Moralities*. New York: Berghahn Books.
- Heintz, M. (2021). *The anthropology of morality: a dynamic and interactionist approach*. Abingdon: Routledge.
- Hongladarom, S. (1999). 'Global Culture, Local Cultures and the Internet: The Thai Example', *AI and Society*, 13(4), pp. 389–401.
- Hongladarom, S. and Ess, C. (2007). *Information Technology Ethics: Cultural Perspectives*. Pennsylvania: IGI Global.
- Hongladarom, S., Joaquin, J.J. and Hoffman, F.J. (2023). 'Global Philosophy of Religion and the Perspectives from Southeast Asia', in Hongladarom, S., Joaquin, J.J. and Hoffman, F.J. (eds.) *Philosophies of Appropriated Religions: Perspectives from Southeast Asia*. Singapore: Springer Nature, pp. 1–8.
- van Hooft, S. (2009). *Cosmopolitanism: A Philosophy for Global Ethics*. Abingdon: Routledge.
- Horáková, N. (2020). 'A Mistrustful Society? The Lack of Trust in Government Institutions in the Czech Republic', *The International Journal of Social Quality*, 10(2), pp. 52–71.
- Van den Hoven, J. and Weckert, J. (eds) (2008). *Information Technology and Moral Philosophy*. Cambridge: Cambridge University Press.
- Jansen, J. (2023). 'When trust becomes a luxury. How economic crises undermine political trust among the most disadvantaged', *Hertie School, Jacques Delors Centre*, 6 June. Available at: <https://www.delorscentre.eu/en/publications/political-trust-in-economic-crisis> (Accessed 12 April 2024).
- Jananoff, S. (2013). 'Epistemic Subsidiarity–Coexistence, Cosmopolitanism, Constitutionalism', *European Journal of Risk Regulation*, 4(2), pp. 133–141.
- Jananoff, S. and Kim, S.-H. (eds) (2015). *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: Chicago University Press.
- Jobin, A., Ienca, M. and Vayena, E. (2019). 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence*, 1(9), pp. 389–399.
- Johnson, D. (2008). *Computer Ethics*. 4th edn. Upper Saddle River: Pearson.
- Kant, I. (1996). 'Groundwork of the Metaphysics of Morals'. Translated by Gregor, M.J. (ed.) *Immanuel Kant, Practical Philosophy, The Cambridge Edition of Immanuel Kant*. Cambridge: Cambridge University Press, pp. 37–108.
- Kant, I. (1996). 'The Metaphysics of Morals'. Translated by Gregor, M.J. (ed.) *Immanuel Kant, Practical Philosophy, The Cambridge Edition of Immanuel Kant*. Cambridge: Cambridge University Press, pp. 353–603.
- Kaun, A., Larsson, A.O. and Masso, A. (2023). 'Automating public administration: citizens' attitudes towards automated decision-making across Estonia, Sweden, and Germany', *Information, Communication & Society*, 27(2), pp. 3141–332.
- Kazim, E. and Koshiyama, A.S. (2021). 'A high-level overview of AI ethics', *Patterns*, 2(9), p. 100314.
- Komaraju, S.A., Arora, P. and Raman, U. (2022). 'Agency and servitude in platform labour: A feminist analysis of blended cultures', *Media, Culture & Society*, 44(4), pp. 672–689.
- Lanzing, M. (2021). 'Contact tracing apps: an ethical roadmap', *Ethics and Information Technology* 23(Suppl 1), pp. 87–90.
- Latour, B. (1994). 'On Technical Mediation', *Common Knowledge*, 3(2), pp. 29–64.

- Levitas, R. (2013). *Utopia as Method*. London: Palgrave Macmillan.
- Lewandowski, D. (2014). 'Why we need an independent index of the Web', in König, R. and Rasch, M. (eds.) *Society of the Query Reader: Reflections on Web Search*. Institute of Network Cultures, pp. 49-58.
- Lewandowski, D. (2019). 'The web is missing an essential part of infrastructure: An Open Web Index', *Communications of the ACM*, 62(4), pp. 24-27.
- Lomborg, S., Kaun, A. and Scott Hansen, S. (2023). 'Automated decision-making: Toward a people-centred approach', *Sociology Compass*, 17(8), p. e13097.
- Lucivero, F. et al. (2022). 'Normative positions towards COVID-19 contact-tracing apps: findings from a large-scale qualitative study in nine European countries', *Critical Public Health*, 32(1), pp. 5-18.
- Van Maanen, G. (2022) 'AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics', *Digital Society*, 1(9).
- Mager, A. (2014). 'Is small really beautiful? Big search and its alternatives', in: in König, R. and Rasch, M. (eds.) *Society of the Query Reader: Reflections on Web Search*. Institute of Network Cultures, pp. 59-72.
- Mager, A. (2017). 'Search engine imaginary: Visions and values in the co-production of search technology and Europe', *Social Studies of Science*, 47(2), pp. 240-262.
- Mager, A. (2018). 'Internet governance as joint effort: (Re)Ordering search engines at the intersection of global and local cultures', *New Media & Society*, 20(10), pp. 3657-3677.
- Mager, A. and Katzenbach, C. (2021). 'Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified', *New Media & Society*, 23(2), pp. 223-236.
- Mager, A. (2023). 'European Search? How to counter-imagine and counteract hegemonic search with European search engine projects', *Big Data & Society* 10(1).
- Mahler, T. (2021). 'Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal', *The Swedish Law and Informatics Research Institute*, 1, pp. 247-270.
- Mayring, P. (2007). 'On Generalization in Qualitatively Oriented Research', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 8(3).
- McLean, K. and Syed, M. (2016). 'Personal, Master, and Alternative Narratives: An Integrative Framework for Understanding Identity Development in Context', *Human Development*, 58(6), pp. 318-349.
- Meuser, M. and Nagel, U. (2009). 'The Expert Interview and Changes in Knowledge Production', in Bogner, A., Littig, B. and Menz W. (eds.) *Interviewing Experts*. Research Methods Series. London: Palgrave Macmillan, pp. 17-42.
- Mladenov, M. (2023). 'Human vs. artificial intelligence - EU's legal response', *Pravo*, 40(1), pp. 32-43.
- Moor, J.H. (1985). 'What Is Computer Ethics?*', *Metaphilosophy*, 16(4), pp. 266-275.
- Munn, L. (2023). 'The uselessness of AI ethics', *AI and Ethics*, 3(3), pp. 869-877.
- Müller, V.C. (2022). 'The History of Digital Ethics', in Véliz, C. (ed.) *Oxford handbook of digital ethics*. Oxford: Oxford University Press, pp.1-19.
- Nagl-Docekal, H. and Zacharasiewicz, W. (eds.) (2022). *Artificial Intelligence and Human Enhancement*. Berlin: de Gruyter.
- OECD (2020). *Good Practice Principles for Data Ethics in the Public Sector*. Available at: <https://www.oecd.org/digital/digital-government/good-practice-principles-for-data-ethics-in-the-public-sector.htm>. (Accessed: 12 April 2024).
- Pagallo, U. (2024). 'Dismantling Four Myths in AI & EU Law Through Legal Information "About" Reality', in Sousa Antunes, H. et al. (eds.) *Multidisciplinary Perspectives on Artificial Intelligence and the Law. Law, Governance and Technology Series*, vol 58. Cham: Springer International Publishing, pp. 251-261.
- Page, R. (2018). *Narratives Online: Shared Stories in Social Media*. Cambridge: Cambridge University Press.
- Palmberger, M. (2022). 'Refugees enacting (digital) citizenship through placemaking and care practices near and far', *Citizenship studies*, 26(6), pp. 781-798.
- Palmberger, M., and Budka, P. (2020). 'Collaborative Ethnography in the Digital Age: Towards a New Methodological Framework', *Digital Ethnography Initiative blog*, 13 November, Available at: <https://digitaletnography.at/2020/11/13/collaborative-ethnography-in-the-digital-age-towards-a-new-methodological-framework/>. (Accessed: 12 April 2024).
- Parker, D.B. (1968). 'Rules of ethics in information processing', *Communications of the ACM*, 11(3), pp. 198-201.
- Pathak-Shelat, M. and Bhatia, K. (2019). 'Examining the significance of online discursive engagement as a form of civic participation', *Global Media Journal, Indian Edition*, 11(1), pp. 2-19.
- Pfotenhauer, S. et al. (2022) 'The politics of scaling', *Social Studies of Science*, 52(1), pp. 3-34.
- Pinch, T. and Collins, H.M. (1982). *Frames of Meaning: The Social Construction of Extraordinary Science*. London: Routledge.
- Pinch, T.J. and Bijker, W.E. (1984). 'The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other', *Social Studies of Science*, 14(3), pp. 399-441.
- Prainsack, B. (2022). 'The Roots of Neglect: Towards a Sociology of Non-Imagination', *Tecnoscienza - Italian Journal of Science & Technology Studies*, 13(2), pp. 13-34.
- Prainsack, B. and Forgó, N. (2024). 'New AI regulation in the EU seeks to reduce risk without assessing public benefit', *Nature Medicine*.
- Prainsack, B. et al. (2022). 'Data solidarity: a blueprint for governing health futures', *The Lancet Digital health*, 4(11), pp. e773-e774.

- Prasad, A. (2014). *Imperial Technoscience: Transnational Histories of MRI in the United States, Britain, and India*. Boston: MIT Press.
- Reijers, W. and Coeckelbergh, M. (2020). *Narrative and Technology Ethics*. Cham: Springer International Publishing.
- Reijers, W. et al. (2018) 'Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations', *Science and engineering ethics*, 24(5), pp. 1437–1481.
- Reviglio, U. and Alunge, R. (2020). "'I Am Datafied Because We Are Datafied": an Ubuntu Perspective on (Relational) Privacy', *Philosophy & Technology*, 33, pp. 595–612.
- Rhodes, C. and Brown, A.D. (2005). 'Narrative, organizations and research', *International Journal of Management Reviews*, 7(3), pp. 167–188.
- Ricourte, P. (2022). 'Ethics for the majority world: AI and the question of violence at scale', *Media, Culture & Society*, 44(4), pp. 726–745.
- Roche, C., Wall, P.J. and Lewis, D. (2023). 'Ethics and diversity in artificial intelligence policies, strategies and initiatives', *AI Ethics*, 3, 1095–1115.
- Le Roux, J. and Oyedemi, T.D. (2022) 'Indelible apartheid: intergenerational post-colonial narratives of colonial-born coloured females about hair, race and identity in South Africa', *Social identities*, 28(2), pp. 150–165.
- Ruttkamp-Bloem, E. (2023). 'Epistemic Just and Dynamic AI Ethics in Africa', in Eke, D.O., Wakunuma, K. and Akintoye, S. (eds.) *Responsible AI in Africa. Social and Cultural Studies of Robots and AI*. Cham: Palgrave Macmillan, pp. 13–34.
- Sardar, Z. (2010). 'The Namesake: Futures; futures studies; futurology; futuristic; foresight—What's in a name?', *Futures*, 42, pp. 177–184.
- Von Schomberg, R. (2013). 'A Vision of Responsible Research and Innovation', in Owen, R., Bessant, J. and Heintz, M. (eds.) *Responsible Innovation*. London: Wiley, pp. 51–74.
- Schütte, B. (2024). 'AI Regulation in the EU - The Future Interplay Between Frameworks', in Gill-Predro, E. and Moberg, A. (eds.) *YSEC Yearbook of Socio-Economic Constitutions 2023: Law and the Governance of Artificial Intelligence*. Cham: Springer Nature, pp. 1–33.
- Silverman, D. (2021). *Doing qualitative research*. Washington: SAGE.
- Spector-Mersel, G. (2010). 'Narrative research: Time for a paradigm', *Narrative Inquiry*, 20(1), pp. 204–224.
- Stiegler, B. (1998). *Technics and Time, 1: The Fault of Epimetheus*. Translated by Collins, G. and Beardsworth, R.. Stanford: Stanford University Press.
- Strathern, M. (2004). *Partial Connections*. Updtd Edn. Walnut Creek: AltaMira.
- Stråth, B. and Wodak, R. (2009). 'Europe - Discourse - Politics - Media - History: Constructing "Crises"?'', in Triandafyllidou, A., Wodak, R. and Krzyżanowski, M. (eds.) *The European Public Sphere and the Media*. London: Palgrave Macmillan, pp. 15–33.
- Suzuki, S. (2023). 'Soft Paternalism and Freedom in the Age of Artificial Intelligence – Through the „tactfulness (融通無碍 Yuzu-Muge)" of 華嚴学 Hua-Yan philosophy', *Paragrana*, 32(1), pp. 247–256.
- Suzuki, S. (2024a). [Forthcoming]. 'Manifesto of Becoming Planetary Theory and the Possibility of Planetary Ethics', *Paragrana*, 32(2).
- Suzuki, S. (2024b). [Forthcoming]. 'The Coevolutionary Path: Planet, AI, and Humanity', *Transition – Journal of the Asian Society of Japan*.
- TallBear, K. (2013). *Native American/ISTA: Tribal Belonging and the False Promise of Genetic Science*. Minneapolis: University of Minnesota Press.
- Taylor, L., Floridi, L. and van der Sloot, B. (eds.) (2017). *Group Privacy: new challenges of data technologies*. Dordrecht: Springer.
- Thompson C.M. (2002). 'Situated knowledge: Feminist and science and technology studies perspectives', in Smelser, N. J., Wright, J. and Baltes, P. B., (eds.) *International Encyclopedia of the Social & Behavioral Sciences*. London: Elsevier, pp.14129–14133.
- Tsing, A.L. (2005). *Friction : an ethnography of global connection*. Princeton: Princeton University Press.
- Umbrello, S. et al. (2023). 'From speculation to reality: Enhancing anticipatory ethics for emerging technologies (ATE) in practice', *Technology in Society*, 74, p. 102325.
- UNESCO (2022). *Recommendation on the Ethics of Artificial Intelligence. Adopted on 23 November 2021*. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (Accessed: 12 April 2024).
- Uruña, S. (2019). 'Understanding "plausibility": A relational approach to the anticipatory heuristics of future scenarios', *Futures*, 111, pp. 15–25.
- Verbeek, P.-P. (2008). 'Obstetric Ultrasound and the Technological Mediation of Morality: A Postphenomenological Analysis', *Human Studies*, 31(1), pp. 11–26.
- Visvanathan, S. (1987). 'From the Annals of the Laboratory State', *Alternatives*, 12(1), pp. 37–59.
- Walsh, T. (2022). *Machines Behaving Badly: The Morality of AI*. Collingwood: La Trobe University Press and Black Inc Books.
- Weckert, J. (ed.) (2019). *Computer Ethics*. London: Routledge.
- Wendehorst, C. and Woopen, C. et al. (2019): *Opinion of the Data Ethics Commission*. Berlin: Data Ethics Commission. <https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/datenethikkommission-abschlussgutachten-lang.html>
- Wendehorst, C. (2021). *The Proposal for an Artificial Intelligence Act COM (2021) 206 from a Consumer Policy Perspective*. Vienna: Federal Ministry of Social Affairs, Health, Care and Consumer Protection. Available at: <https://www.sozialministerium.at/dam/sozialministe>

- riumat/Anlagen/Themen/Konsumentenschutz/Konsumentenpolitik/The-Proposal-for-an-Artificial-Intelligence-Act-COM2021-206-from-a-Consumer-Policy-Perspective_dec2021__pdfUA.pdf. (Accessed: 12 April 2024).
- Wendehorst, C. (2023). 'Konsument:innen in der Datenökonomie', in Raiffenstein, M. Blaschek, B. (eds.) *Konsumentenpolitisches Jahrbuch 2023 – Trends, Rechtsentwicklung und Judikatur der letzten zwei Jahre*. Wien: Verlag Österreich, pp. 215–253.
- White, H. (1987). *The Content of the Form Narrative Discourse and Historical Representation*. Baltimore: The Johns Hopkins University Press.
- WHO (2021) *Ethics and Governance of Artificial Intelligence for Health. WHO Guidance*. Available at: <https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1>. (Accessed: 12 April 2024).
- Wittmayer, J.M. et al. (2019) 'Narratives of change: How social innovation initiatives construct societal transformation', *Futures*, 112, p. 102433.
- Wodak, R., Reisigl, M. and de Cillia, R. (2022). *The Discursive Construction of National Identity*. Edinburgh: Edinburgh University Press.
- Wong, P. (2012). 'Dao, harmony and personhood: towards a confucian ethics of technology', *Philosophy & Technology*, 25, pp. 67–86.
- Woolgar, S. (2002). 'Five Rules of Virtuality', in Woolgar, S. (ed.) *Virtual society?: Technology, cyberbole, reality*. Oxford: Oxford University Press, pp. 1–22.
- Zhang, J.Y. and Datta Burton, S. (2022). *The elephant and the dragon in contemporary life sciences: a call for decolonising global governance*. Manchester: Manchester University Press.

LIST OF FIGURES

Figure 1: Methodology of the AGIDE project	16
Figure 2: Matrix of Digital Ethics Narratives	40
Figure 3: "Coloniality"-type pattern	43
Figure 4: Variant of the "Coloniality"-type pattern.....	44
Figure 5: "Beneficiary" pattern.....	44
Figure 6: "Beneficiary" pattern turning into "Actor" pattern	45
Figure 7: Japan-associated variant of the "Harmony-Opportunity"-type pattern	45
Figure 8: China-associated variant of the "Harmony-Opportunity"-type pattern	46
Figure 9: Recent developments with regard to the China-associated variant of the "Harmony-Opportunity" narrative pattern	46
Figure 10: Original "Silicon Valley-type approach.....	47
Figure 11: Moving away from the traditional "Silicon Valley" pattern?	47
Figure 12: A typical (more secular) Israel-associated pattern	48
Figure 13: The traditional "GDPR"-type pattern	48
Figure 14: Traditional approach in the "Nordic countries"	49
Figure 15: Denmark after the "Chromebook" and other scandals	50
Figure 16: A government-sceptic pattern from CEE countries.....	50
Figure 17: Post-Brexit UK – moving from the "GDPR" to the "Silicon Valley" pattern?.....	51
Figure 18: Risk-based approach of the AI Act – few changes required	51
Figure 19: Data economy legislation – change of the law but not of the narrative	52
Figure 20: Hypothetical "EU digital sovereignty" narrative..	53
Figure 21: India – upwardly mobile youth.....	53
Figure 22: India – older generation.....	54
Figure 23: India – official government discourse.....	54
Figure 24: India – public interest groups and intellectuals	55
Figure 25: Spheres of factors influencing digital ethics narratives [source: Working Group Members]	58

ANNEX: MEMBERS OF THE AGIDE WORKING PARTY



AKUETTEH, Teki is an ICT/telecom lawyer, a privacy consultant and senior partner at a law firm based in Accra, Ghana. She is also the founder and executive director of the Africa Digital Rights Hub, and a non-resident fellow of the Center for Global Development. Previously, Teki worked for the

government of Ghana in the development of several key legislations for the ICT sector. She was also the first executive director of the Data Protection Commission of Ghana and a member of the UN Global Pulse Privacy Advisory Group. Teki holds an LLM in IT and Telecommunications Law from the University of Strathclyde, Glasgow, Scotland and a Bachelor of Arts in Law and Political Science from the University of Ghana (Legon), Accra, Ghana.



ALMEIDA, Virgilio is a professor emeritus of Computer Science at the Federal University of Minas Gerais (UFMG). He is also faculty associate at the Berkman Klein Center at Harvard University. He held visiting positions in several universities and research labs, such as Harvard University (School of Engineering and

Applied Sciences), New York University, Boston University, Santa Fe Institute, HP Labs. Almeida is co-author of five books covering topics such as web technologies, e-commerce, performance modelling, and capacity planning, published by Prentice Hall. He is also the author of *Governance for the Digital World*, published by Palgrave MacMillan, and his most recent book is *Algorithmic Institutionalism: The Changing Rules of Social and Political Life*, published by Oxford University Press. Almeida served as one of the commissioners of the Global Commission for the Stability of Cyberspace. His current research interests revolve around social computing, governance of algorithms, the social impact of AI, and modelling and analysis of large-scale distributed systems.



AMEEN, Nisreen is a senior lecturer (associate professor) in Digital Marketing at Royal Holloway, University of London and director of the Digital Organization and Society research centre. Nisreen is also currently serving as vice-president of the UK Academy of Information Systems (UKAIS).

She is an associate editor for *Information Technology and People* and *Computers in Human Behavior*. Nisreen has also served as a guest editor for special issues of various top-ranked journals. Nisreen's research focuses on two main themes: first, consumer interactions with new-age technologies, and second, how organisations can use these technologies to provide better experiences, services and products. Within these main themes, her research interests include digital marketing, human-computer interaction, consumer behaviour, and organisational use of consumer data. She also focuses on cross-national and cross-cultural research in developing, emerging and developed markets. Nisreen's work has been featured in world-leading media outlets such as CNN, World Bank, The Conversation, New Statesman and Fast Company.



ARORA, Payal is a professor at Utrecht University, and a co-founder of FemLab, a feminist future of work initiative. She is a digital anthropologist, TEDx speaker, and author of award-winning books, including *The Next Billion Users* with Harvard Press. Her expertise lies in user experience in the

Global South and inclusive design. Several international media outlets have covered her work, including the BBC, *The Economist*, Quartz, *Tech Crunch*, *The Boston Globe*, *F.A.Z.*, *The Nation*, and the CBC. Forbes named her the "next billion champion" and the "right kind of person to reform tech". She has consulted for numerous organisations including

Adobe, UNHCR, GE, HP, Interpol, UNDP, Spotify, and many others. Her upcoming book is with MIT Press titled *From Pessimism to Promise: Lessons from the Global South on building Inclusive Tech*.



ASPELMEYER, Markus (OeAW) is professor of Physics at the University of Vienna and scientific director at the Institute for Quantum Optics and Quantum Information (IQOQI) of the Austrian Academy of Sciences in Vienna. Aspelmeyer was born in the Bavarian town Schongau, where he also

attended the local school. He studied physics and philosophy in Munich, Germany. Aspelmeyer is regarded as one of the pioneers of the field of quantum optomechanics. His research combines the development of new quantum technologies with fundamental quantum experiments. He is co-founder of Crystalline Mirror Solutions (now Thorlabs Crystalline Solutions), which provides novel optics for laser precision measurements. He is a fellow of the American Physical Society, and member of the Austrian Academy of Sciences and the Academy of Sciences and Humanities in Hamburg. His current research is focused on the intriguing puzzles around quantum physics and gravity.



BEN-DAVID, Anat is an associate professor at the Department of Sociology, Political Science and Communication at the Open University of Israel. She is a co-founder of the Open University's Open Media and Information Lab (OMILab). Focusing on new media, her primary research interests are

the history and geopolitics of the web, Digital STS, social and political studies of social media, and digital and computational methods for web research.



BROWN, Keith A. (NAS) is an associate professor of Mechanical Engineering, Materials Science & Engineering, and Physics at Boston University. He completed undergraduate studies at the Massachusetts Institute of Technology, doctoral studies at Harvard University, and postdoctoral studies at North-

western University. His research group explores approaches to accelerate the development of advanced materials and

structures using self-driving labs. Keith has co-authored over 100 peer-reviewed publications and has six issued patents. Keith has received the Frontiers of Materials Award from The Minerals, Metals, & Materials Society (TMS), been named a "Future Star of the AVS", and received the Omar Farha Award for Research Leadership from Northwestern University. Keith served on the Nano Letters Early Career Advisory Board, co-organised a National Academies of Science, Engineering, and Medicine Workshop on AI for Scientific Discovery, and currently leads the MRS Artificial Intelligence in Materials Development Staging Task Force.



CHIN, Yik Chan is an associate professor in the School of Journalism and Communication at Beijing Normal University. She also previously worked at the University of Nottingham and the University of Oxford's School of Law. Her research interests include internet governance, digital ethics, policy,

regulation and law, and AI and data governance. Her ongoing projects include digital governance in China, global AI and data governance. She has authored a wide range of refereed journal articles, book chapters, research monographs and policy reports. She is the author of *Television Regulation and Media Policy in China* (Routledge, 2017) and her upcoming monograph is on *Governance of the Digital in China* (2025). Her latest publication includes *Right to Data Access in the Digital Era: the Case of China* (2024). Dr. Chin is the vice-chair of the Public Media Policies Working Group of IAMCR, a member of the UN IGF Policy Network on Artificial Intelligence (PNAI), NGO representative to the UN's OEWG process, a member of the Multi-Stakeholder Steering Group of Asian Pacific Internet Governance Forum (APrIGF) and a Multi-Stakeholder Steering Group member of the China Internet Governance Forum (China IGF).



EITENBERGER, Magdalena is a postdoctoral researcher at the Centre for the Study of Contemporary Solidarity (CeSCoS). Her work focuses on issues of digital ethics and health, social policy, and healthcare technologies and systems. Magdalena teaches on health policy, gender in medicine and health, medical

ethics, and research methods at the University of Vienna, and on problem-oriented learning and public health at the Medical University of Vienna. Her PhD project, titled „Technologizing Care in Chronic Illness“, explored type 1 diabetes technologies and policy decision-making in

Austria. Magdalena earned a PhD in the Department of Political Science at the University of Vienna, a Master's degree in Science & Technology Studies at the University of Vienna and Maastricht University, a Master's degree in Communications, and a Bachelor's degree in Political Science from the University of Vienna. Before joining the Department of Political Science (2023), Magdalena worked as a postgraduate researcher at the Ludwig Boltzmann Institute for Digital Health and Patient Safety and the Institute for Ethics and Law in Medicine, and as a project manager at the Federal Ministry of Social Affairs, Health Care, and Consumer Protection during the COVID-19 pandemic.



HALLER, Max (OeAW) is Prof. em. of Sociology and vice-chair of the Commission for Migration and Integration Research at the OEAW. Max Haller, born 1947 in Sterzing (Italy), 1985–2015 Prof. of Sociology at the University of Graz. He was president of the Austrian Sociological Association and is a

member of the Austrian Academy of Sciences. Co-founder and vice-president of the European Sociological Association and co-founder of the International Social Survey Programme (ISSP). He was a visiting professor at universities in Austria, Germany, Hungary, the Czech Republic, Trento (Italy), Sta. Barbara (California, USA) and the St. Augustine University of Tanzania (Tanzania). He published and edited 40 books and about 250 papers in international sociological Journals (including *AJS*, *ASR*, *Review Francaise de Sociology*, *International Sociology*) and readers. His main research areas are social stratification, sociology of European integration, comparative social research, etc.



HONGLADAROM, Soraj is a research fellow at the Center for Science, Technology and Society at Chulalongkorn University in Bangkok, Thailand. He has published books and articles on such diverse issues as bioethics, computer ethics, and the roles that science and technology play in the culture of developing countries.

A large part of his research focuses on how information technology is integrated in the lifeworld of the Thai people. Books authored by him include *The Ethics of AI and Robotics: A Buddhist Viewpoint* (Rowman and Littlefield), *The Online Self* (Springer) and *A Buddhist Theory of Privacy* (Springer). His articles have appeared in *The Information Society*, *AI & Society*, *Philosophy in the Contemporary World*, and *Social Epistemology*, among others.



JEMIELNIAK, Dariusz (PAN) is a full professor and head of Management at the Networked and Digital Environments (MINDS) department, Kozminski University, and a faculty associate at the Berkman-Klein Center for Internet and Society, Harvard University. He is a member of the Polish

Academy of Sciences and its current vice-president. His recent books include *Strategizing AI In Business and Education* (2023, Cambridge University Press, with A. Przegalinska), *Collaborative Society* (2020, MIT Press, with A. Przegalinska), *Thick Big Data* (2020, Oxford University Press, and *Common Knowledge? An Ethnography of Wikipedia* (2014, Stanford University Press). His current research projects include disinformation and bot detection. He currently serves on the Wikimedia Foundation Board of Trustees.



KEEGAN, Te Taka is from the Waikato-Maniapoto, Ngāti Porou and Ngāti Whakaue tribes. He is an associate professor of Computer Science and the associate dean Māori of the Wānanga Pūtaiao at the University of Waikato. He has a Diploma in Computer Engineering from CIT, Wellington

(1987) and a PhD (2007) from Waikato University. Te Taka has worked on a number of projects involving the Māori language and technology. These include the Māori Niupepa Collection, Te Kete Ipurangi, the Microsoft keyboard, Microsoft Windows and Microsoft Office in Māori, Moodle in Māori, Google Web Search in Māori, the Māori macroniser and SwiftKey for Māori. In 2013, Te Taka was awarded the University of Waikato's Māori/Indigenous Excellence Award for Research. In 2017, Te Taka was awarded the Prime Minister's Supreme Award for Tertiary Teaching Excellence.



LIGETT, Katrina (IASH) is a professor of Computer Science and director of the Federmann Center for the Study of Rationality at the Hebrew University. Her research deals with data privacy, algorithmic fairness, and algorithmic game theory. In addition to her contributions to basic research, her work

seeks out connections with other fields, including policy, law, and ethics. Katrina has won a large number of research

grants, including various grants from the American NSF, Israeli ISF, and European ERC. Her articles are published in the leading journals and conferences in her fields, and she is frequently invited to lecture in academia, industry and the public sector.



MAGER, Astrid (OeAW) is a senior academy scientist at the Institute of Technology Assessment (ITA), Austrian Academy of Sciences, and lecturer at the Department of Science and Technology Studies (STS), University of Vienna. Her long-standing research on search engines/algorithmic systems in socio-political contexts has been widely published.

Recent publications include *European Search? How to counter-Imagine and counteract hegemonic search with European search engine projects*, *Big Data & Society* (2023), *Advancing search engine studies: The evolution of Google critique and intervention*, *Big Data & Society* (2023, together with O. C. Norocel and R. Rogers), and *Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified*, *New Media & Society* (2021, together with C. Katzenbach). In 2018, Mager was elected to the Young Academy of the Austrian Academy of Sciences.



MAO, Yishu is a PhD candidate at the Department of Communication and Media Research, University of Zurich, and a predoctoral fellow in the RG "China in the Global System of Science" at the Max Planck Institute for the History of Science. Her PhD project analyses emerging AI governance regimes in China, Germany and the US by employing comparative policy analysis and STS methods. She seeks to identify how politics, academia, and industry in the three countries shape the understanding of how to govern AI, and how they diverge and converge. Her broad research interests include the co-production of science and social order, communicative construction of AI imaginaries, and global AI governance. She worked at the Mercator Institute for China Studies (MERICS) and the Center on U.S.–China Relations at the Asia Society on a range of topics including the ideology of Chinese overseas students, digital development, and the intersection of online public opinion and policymaking in China. She holds an MA in Global Studies from the Humboldt University of Berlin and a BA in Literature and Digital Humanities from Bard College.



MOODLEY, Keymanthri is a distinguished professor at the Department of Medicine and head of the Division of Medical Ethics and Law, Faculty of Health Sciences, Stellenbosch University. The division was re-designated as a Collaborating Centre in Bioethics in 2023 for the third time by the World

Health Organization (WHO), becoming one of twelve in the world and the first on the African continent. Keymanthri is also an adjunct professor at the Department of Social Medicine, University of North Carolina at Chapel Hill, USA. She is a specialist family physician, a bioethicist, an NRF-rated researcher (established scholar with international recognition in 2019) and a member of the Academy of Science of South Africa (ASSAf). To date, she has been awarded grants to the approximate value of R70 million. The division has trained 40 mid-career professionals from 10 African countries in bioethics and currently has 4 PhD graduates and 4 PhD trainees. Keymanthri serves on the Scientific Advisory Committee of EDCTP. Her most recent NIH-funded research project explores ethical, legal, and social issues in data science innovation in Africa.



NAGL-DOCEKAL, Herta (OeAW) is Prof. of Philosophy (i.R.), University of Vienna, Austria; Dr. h. c. Univ. of Basel, Switzerland; full member, Austrian Academy of Sciences; membre tit., Institut International de Philosophie; vice-president of FISP (2008–2013). Visiting prof.: Utrecht, the Netherlands; Free

Univ. Berlin, Univ. of Konstanz, and J.W. Goethe Univ., Frankfurt am Main, Germany, University of St. Petersburg, Russian Federation. Selected books: *Religion in the Secular Age* (co-ed. 2023), *Artificial Intelligence and Human Enhancement. Affirmative and Critical Approaches in the Humanities* (co-ed., Berlin 2022), *Leibniz heute lesen* (ed., Berlin 2018), *Innere Freiheit. Grenzen der nachmetaphysischen Moralkonzeptionen* (Berlin 2014), *Glauben und Wissen. Ein Symposium mit Jürgen Habermas* (co-ed., Berlin 2007), *Feminist Philosophy* (Cambridge, MA, 2004), *Continental Philosophy in Feminist Perspective* (co-ed., University Park 2000), *Postkoloniales Philosophieren: Afrika* (co-ed., Vienna 1992). Co-editor *Deutsche Zf. für Philosophie* (1993–2005).



NATARAJAN, Sarayu is the founder of the Aapti Institute. Aapti works for equity and justice in technology for individuals and communities and specialises in research, action, and investment concerning inclusion, governance, and data through its two labs, the Digital Public Lab and the Data Econ-

omy Lab. Sarayu has a PhD in Political Science from King's College London, an MPA from the School of International and Public Affairs at Columbia University, and a law and arts degree from the National Law School of India University, Bangalore. She has previously also worked at McKinsey and Company, Elevor Equity, and Gray Matters Capital.



OLIVEN, Ruben George was director (2019–2022) and is currently one of the vice-presidents of the Brazilian Academy of Sciences (2022–2025). Professor of the Postgraduate Program in Social Anthropology of the Federal University of Rio Grande do Sul (UFRGS), Ruben Oliven has a Bachelor in Social

Sciences and Economic Sciences, Master's in Urban Planning by UFRGS and PhD in Social Sciences from the University of London. He is a researcher at CNPq (Brazil's Science Foundation), and was a visiting professor at foreign universities, including the University of California (Berkeley), Brown University, the University of Illinois, Emory University, the University of Paris and Leiden University. He is past president of the Brazilian Association of Anthropology (ABA) and of the National Postgraduate and Social Sciences Association (ANPOCS). He received the Erico Vannucci Mendes award for his contribution to the study of Brazilian culture and the ANPOCS Award of Academic Excellence in Anthropology. In 2018, he was awarded the Grand Cross of the National Order of Scientific Merit, the highest distinction granted to Brazilian scientists.



PATHAK-SHELAT, Manisha is professor of Communication & Digital Platforms and Strategies at MICA, Ahmedabad, India and co-chairs MICA's Centre for Development Management and Communication. She is the editor of the *Journal of Creative Communications* published by Sage with MICA. She has a

Ph.D. in Mass Communication with a minor in Gender and

Women's Studies from the University of Wisconsin-Madison, USA and in Education from the Maharaja Sayajirao University of Baroda, India. Manisha's research interests are media and information literacy, digital cultures, inclusive design, communication for social change, gender, and transcultural citizenship. Awards and fellowships include the Soviet Land Nehru Award, the Shastri Indo-Canadian Faculty Research Award, the Salzburg Seminar Fellowship, the TATA Fellowship for the Study of Contemporary India, the MICA AGK Award for meritorious service, the lifetime achievement award for contribution to media and communication education of the Global Media Education Council, and the Sharon Dunwoody Award from the University of Wisconsin-Madison School of Journalism and Mass Communication for outstanding contribution to the field.



PRAINSACK, Barbara (OeAW) is a professor at the Department of Political Science at the University of Vienna, where she directs the Centre for the Study of Contemporary Solidarity (CeSCoS). Her latest books are: *The Pandemic Within: Policy Making for a Better World* (with H. Wagenaar, Policy Press, 2021);

and *Personalised Medicine: Empowered Patients in the 21st Century?* (New York University Press, 2017). Barbara is also a member of the Austrian National Bioethics Committee, and chair of the European Group on Ethics in Science and New Technologies advising the European Commission. She holds an honorary professorship at the School of Social and Political Sciences at the University of Sydney and is an elected foreign member of the Danish Royal Academy of Sciences and Letters, an elected member of the Academia Europaea (AE), an elected member of the German National Academy of Science and Engineering (acatech), and an elected corresponding member of the Austrian Academy of Sciences. In the academic year 2023/24, she is a fellow at the Institute of Advanced Study (Wissenschaftskolleg) in Berlin.



RUTTKAMP-BLOEM, Emma is a philosopher of science and technology, an AI ethics policy adviser, and a machine ethics researcher. She is the head of the Department of Philosophy at the University of Pretoria, the current chair of the UNESCO World Commission on the Ethics of Scientific Knowledge and

Technology (COMEST), and a member of the UN Secretary-General's High Level Advisory Body on AI. She is the

AI ethics lead at the Centre for AI Research (CAIR), and the chair of the Southern African Conference on AI Research (SACAIR). She was the chairperson of the UNESCO Ad Hoc Expert Group that prepared the draft of the 2021 *UNESCO Recommendation on the Ethics of AI*. Emma is a member of the African Commission Human and Peoples' Rights Committee (ACHPR) task team working on the Resolution 473 study on Human and Peoples' Rights and AI, Robotics and other New and Emerging Technologies in Africa. She is a member of the Global Academic Network at the Center for AI and Digital Policy, Washington, DC. She sits on various international AI ethics advisory boards and consults widely in the private sector.



SCHMIDT, Albrecht is a professor of Computer Science at Ludwig-Maximilians-Universität (LMU) in Munich, where he holds a chair for Human-Centered Ubiquitous Media. His research interests are in intelligent interactive systems, ubiquitous computing, digital media technologies and media informatics. He studied Computer Science in Ulm and Manchester and in 2003 he completed his PhD at Lancaster University. In 2018, he was inducted into the ACM SIGCHI Academy, and in 2020, he was elected to the Leopoldina, the Germany academy of science, being named ACM Fellow in 2023.

He studied Computer Science in Ulm and Manchester and in 2003 he completed his PhD at Lancaster University. In 2018, he was inducted into the ACM SIGCHI Academy, and in 2020, he was elected to the Leopoldina, the Germany academy of science, being named ACM Fellow in 2023.



SHINWARI, Zabta Khan, a UNESCO laureate and distinguished national professor, has made significant contributions to academia, bioethics, and sustainable development in Pakistan. He holds a PhD from Kyoto University, Japan, and has held esteemed positions at various institutions including

Quaid-i-Azam University and Kohat University of Science & Technology. Dr. Shinwari's research in molecular systematics & medicinal plants, bioethics, and biotechnology has led to breakthroughs in stress tolerance genes and extensive scholarly output, including 8 authored books, 14 edited international proceedings, and over 530 articles in impact factor journals. With an impressive citation record of over 21550, mentoring over 100 M.Phil & PhD students, and securing numerous grants, Dr. Shinwari is a distinguished figure in scientific circles. His accolades include civil awards from the Government of Pakistan, recognition

by UNESCO, and prestigious titles from scientific academies worldwide. Dr. Shinwari's focus on biosecurity education, ethics in frontier technologies and sustainable development underscores his commitment to fostering ethical practices and economic growth in Pakistan's biotechnology sector, further evidenced by his collaborations with international agencies and advocacy for open science principles.



SUZUKI, Shoko PhD is a professor emeritus at Kyoto University, a principal researcher at RIKEN and a visiting researcher at the Information and Communications Policy Institute, Ministry of Internal Affairs and Communications in Japan. Specialised in the philosophy of science and anthropology, she

was a professor at Kyoto University from 2003 to 2022, a member of the Science Council of Japan from 2005 to 2020, a visiting professor at the Free University of Berlin, Germany, 2009–2010, Excellence Graduate School Program „Language of Emotion“. Since 2016, team leader, Artificial Intelligence Ethics and Society Team, RIKEN Centre for Advanced Intelligence Project (AIP); 2016, member of the Roundtable on Artificial Intelligence and Society, Cabinet Office; 2016, secretary, AI Network Society Promotion Council, Ministry of Internal Affairs and Communications. In 2018, she received a Ministry of Education, Culture, Sports, Science and Technology Award for Distinguished Contributions to Local Administration 2019.



TAYLOR, Linnet is professor of International Data Governance at the Tilburg Institute for Law, Technology, and Society (TILT) in the Netherlands. Her research focuses on digital data, representation and democracy, with particular attention to transnational governance issues. Her work on

group privacy and data justice is used in discussions of technology governance in countries around the world. She also led the Global Data Justice project (2018–23), aiming to develop a social-justice-informed framework for governance of data technologies on the global level. Her work has also been supported by the Luminare foundation and the EU AI Fund. She is a member of the Dutch Young Academy (De Jonge Akademie).



WALSH, Toby is the chief scientist of UNSW.AI, UNSW's new AI Institute. He is a strong advocate for limits to ensure AI is used to improve our lives, having spoken at the UN and to heads of state, parliamentary bodies, company boards and many others on this topic.

This advocacy has led to him being „banned indefinitely“ from Russia. He is a fellow of the Australian Academy of Science, and was named on the international „Who's Who in AI“ list of influencers. He has written four books on AI for a general audience, the most recent is *Faking It! Artificial Intelligence in A Human World*.



WEISS, Astrid (OeAW) is assistant professor on Human Interactions with Embodied AI at the TU Wien, Austria. With a background in sociology and human-computer interaction (HCI), she studies how humans interact with new technology with a special interest in user involvement and evaluation

studies for human-robot interaction (HRI), focusing on in-the-wild studies and controlled experiments. She frequently works with engineers, computer scientists, interaction designers, psychologists, and representatives from other related disciplines to shape a technology development process that creates interactive systems people sustainably use in their everyday environments and contexts.



WENDEHORST, Christiane (OeAW) is president of the Humanities and Social Sciences Division of the Austrian Academy of Sciences. She has been professor of Law at the University of Vienna since 2008 and is, among other things, a founding member, past president (2017–2021) and since 2022 scientific

director of the European Law Institute (ELI). She is also an elected member of the Academia Europea, the International Academy for Comparative Law (IACL) and the American

Law Institute (ALI). Before moving to Vienna, she held professorships in Göttingen (1999–2008) and Greifswald (1998–1999) and was director of the Sino-German Institute of Legal Studies (2000–2008). Wendehorst's current research focuses on the legal challenges of digitalisation, and she regularly advises EU institutions and national governments on matters relating to law and regulation in the digital economy.



WINTER, Jana is the project coordinator and research associate for AGIDE „Academies for Global Innovation and Digital Ethics“ at the Austrian Academy of Sciences (AAS). She is currently writing her Master's thesis on values, design and freedom under the supervision of Prof. Coeckelbergh at the

Department of Philosophy (University of Vienna). During her undergraduate studies, she worked as a tutor for Ass. Prof. Romizi, Ass. Prof. Graness and Univ. Prof. Stenger (Ethics, Intercultural Philosophy, Political Philosophy, Philosophy in a Global World). In 2020, she worked as a research assistant to Ass. Prof. Weng on an interdisciplinary project involving engineering, law and ethics, focusing on human-robot interaction in Japan (Tohoku University), which sparked her interest to graduate in Interdisciplinary Ethics with a special focus on Digital Ethics and Philosophy of Technology.



ZENG, Yi (CAS) is a professor and director of the Brain-inspired Cognitive Intelligence Lab, and founding director of the International Research Center for AI Ethics and Governance, both at the Institute of Automation of the Chinese Academy of Sciences. He is a founding director of the AI for

SDGs Cooperation Network and the Center for Long-term AI. He is a member of the National Governance Committee of the New Generation AI, China. He is also a member of the Advisory Council for the Institute of Ethics in AI, University of Oxford, a member of the UNESCO Ad Hoc Expert Group on AI Ethics, and a member of the WHO Expert Group on AI Ethics and Governance for Healthcare.



ISBN 978-3-7001-9705-8

—
ÖAW
—

WWW.OEAW.AC.AT