Ethical framework for artificial intelligence and digital technologies

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.ijinfomgt.2021.102433

Publisher: Elsevier

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Ethical Framework for Artificial Intelligence and Digital Technologies

Abstract

The use of Artificial Intelligence (AI) in Digital technologies (DT) is proliferating a profound socio-technical transformation. Governments and AI scholarship have endorsed key AI principles but lack direction at the implementation level. Through a systematic literature review of 59 papers, this paper contributes to the critical debate on the ethical use of AI in DTs beyond high-level AI principles. To our knowledge, this is the first paper that identifies 14 digital ethics implications for the use of AI in seven DT archetypes using a novel ontological framework (physical, cognitive, information, and governance). The paper presents key findings of the review and a conceptual model with twelve propositions highlighting the impact of digital ethics implications on societal impact, as moderated by DT archetypes and mediated by organisational impact. The implications of intelligibility, accountability, fairness, and autonomy (under the cognitive domain), and privacy (under the information domain) are the most widely discussed in our sample. Furthermore, ethical implications related to the governance domain are shown to be generally applicable for most DT archetypes. Implications under the physical domain are less prominent when it comes to AI diffusion with one exception (safety). The key findings and resulting conceptual model have academic and professional implications.

Keywords: Artificial Intelligence (AI) ethics, digital ethics, digital technologies and archetypes, PRISMA, systematic literature review, ontological framework

1. Introduction

Processes, systems, and supply chains that drive our daily lives, from energy grids to healthcare, food distribution to online banking, are increasingly dependent on digital technologies (DT). The physical world is being digitised and the "second machine age" (Brynjolfsson & McAfee, 2014, p. 7) characterised by the use of Artificial Intelligence (AI) in DTs has the potential to significantly change society. AI enables machines to execute cognitive tasks with minimal or no human interaction. At the same time, it has a profound impact on organisations by challenging the status quo but promising new value propositions and enhanced (digital) revenue streams (Ashok, 2018; Pathak et al., 2020). Such digital capabilities have been fundamental in weathering social and economic turbulences such as natural disasters, financial crises, pandemics, etc. However, the use of AI in DTs forces us to grapple with tensions between technological and societal advancement on one hand and the same technologies introducing ethical dilemmas on the other hand (Floridi, 2018).

AI development has typically been the purview of a select group of engineers, scientists, programmers, and architects, who have failed to sufficiently represent the ethnic, cultural, gender, age, geographic, or economic diversity of human social life. As systems become fully autonomous, there is a risk humans will become dependent on these systems. When AI is biased by design, humans forfeit their agency to serve the needs of AI rather than its original purpose of serving humans (Applin, 2017). Some recent examples of digital ethical issues include the unjust acquisition of millions of Facebook users data by Cambridge Analytica (Venturini & Rogers, 2019); use of social media and Twitter bots to influence votes outcome in Brexit and 2016 US presidential elections (Gorodnichenko et al., 2021); impact of misinformation and fake news, aggravated by the use of social media, on violence, lynching, riots in India (Khan et al., 2019); an

explosion of cybercrime-related to data hacking and breaches, affecting millions of consumers (personal, payment, behavioural data) (Janakiraman et al., 2018); bias against certain groups of gender, ethnicity, race, etc. by facial analysis software and algorithms (Khalil et al., 2020); gender-based discrimination (by algorithms) in the online display of STEM career advertisements (Lambrecht & Tucker, 2019); and, use of cryptocurrency in fraud and malware (Conti et al., 2018). Thus, we frame the research question as:

What are the key digital ethics implications for the use of AI in digital technologies?

Following the introduction, the paper outlines motivations for the paper and a four-dimensional ontological framework. This is followed by a theoretical discussion on digital ethics and DTs within the context of the ontological framework as an a priori template. Through a systematic literature review and qualitative synthesis, we inductively develop a typology of ethical implications with the use of AI in DTs. Finally, we develop a conceptual model with twelve propositions.

2. Theoretical Framework

2.1 Motivation

Dwivedi et al. (2021a) outline a lack of agreement on the concept of artificial intelligence and suggest that researchers develop discipline-oriented typologies to enable boundary research in social sciences. Similarly, Stahl et al. (2016)'s review of computing ethics literature reveals ethics discussions are broad and generic and need to be focussed on particular technologies to have practical importance. To this effect, the last two years have seen several systematic reviews on AI literature within discipline-oriented streams such as business strategy, sustainability, decision making, digital marketing and social media, internet of things, blockchain, education, and learning (Borges et al., 2020; Duan et al., 2019; Dwivedi et al., 2021b; Kankanhalli et al., 2019; Nishant et

al., 2020; Sarker et al., 2019). To the best of our knowledge, a systematic literature review on the ethical dimensions of the use of AI in DTs is lacking, despite the significant impact of misuse of AI in DTs.

Floridi (2018, p. 2) argues that current AI scholarship lacks a "socio-political direction" and requires a collective vision on the future of AI to be ahead of the pace of technological development. Jobin et al. (2019)'s systematic review of 84 AI guidelines shows a convergence of high-level AI principles but divergence on interpretation and application of these principles. Similarly, Whittlestone et al. (2019, p. 11) state that AI guidelines "are not specific enough to be action-guiding." J. Morley et al. (2020) concurs that AI principles need to be translated to 'what' and 'how' of implementation. Adomavicius et al. (2007) argue that a single technology cannot be considered in isolation and it is essential to evaluate the diversity of different DTs when assessing their ethical impact and implications. There have been attempts to develop implementation-oriented analytical tools, for example, Canada's Algorithmic Impact Assessment (Canada, 2020), World Economic Forum's AI Procurement in a Box (Forum, 2020), OECD's Framework on AI Strategies (Berryhill et al., 2019). However, these lack a common ontological basis risking ethical assessments and corresponding AI development diverging between actors driven by political and economic interpretations of high-level AI principles.

2.2   *Ontological framework*

We adopt Gruber (1993, p. 199)'s definition of ontology as "an explicit specification of a conceptualization." As opposed to the philosophical concept of ontology as the nature of reality, the use of ontology in information science is geared towards defining a "shared taxonomy of entities" (Smith & Welty, 2001, p. vi).

We build on several ontological frameworks (Ogden & Richards, 1923; Popper, 1979; Project & Peirce, 1998) and identify three domains to conceptually understand the ethical impact and interrelationships inherent in the use of AI in DTs as shown in Table 1. The physical domain represents Popper's World I ontology and referent/object in semiotics; the cognitive domain represents Popper's World II and reference/interpretant in semiotics; and, the information domain represents Popper's World III and symbol/science in semiotics (Ibid.). Denning and Rosenbloom (2009) and Rosenbloom (2012) refer to these triadic domains as physical, social, and life respectively. Building on Liu (2000)'s concept of organisational onion, we propose the fourth domain as governance capturing the in(formal) information system and contending technical information system is already captured by the other three domains.

[Table 1: Ontological frameworks for information science]

2.3   Key concepts

2.3.1   Digital Ethics

Fieser (2000) discusses three main schools of thought on ethics: metaethics, normative ethics, and applied ethics. Applied ethics has emerged as a school of thought that combines consequential and nonconsequential approaches in specific contexts such as medical and business ethics (Breidbach & Maglio, 2020; Fieser, 2000). Phillip (1985, p. 381) defines business ethics as "rules, standards, codes, or principles, which provide guidelines for morally right behaviour and truthfulness in specific situations." We expand on business ethics to define digital ethics as "the systems of values and moral principles for the conduct of electronic interactions" (Buytendijk, 2019). Within the context of AI use, these electronic interactions encapsulate both human-machine and machine-machine interactions.

This study views digital ethics as being socially constructed by technological (Utilitarianism perspective) and societal moral values (Kantianism perspective), which act upon each other in a triadic recursive way. Furthermore, the digital ethics space covers both hard ethics, right and wrong actions based on regulations and compliance, and soft ethics, "what ought and ought not to be done over and above the existing regulation" (Floridi, 2018, pp. 4-5).

The literature discusses several context-specific applied ethics research domains related to digital ethics as shown in Table 2. We map these ethical research domains to the four dimensions of our ontological framework and view digital ethics as the focal point of the four domains as shown in Figure 1.

[Table 2: Research domains in applied ethics in digital technologies]

[Figure 1: Digital ethics research domain]

2.3.2   Artificial Intelligence and Digital Technologies (DT) Archetypes

Bain (1937, p. 860)'s definition of technology is widely accepted among social scientists as "all tools, machines, utensils, weapons, instruments, housing, clothing, communicating and transporting devices and the skills by which we produce and use them." Arthur (2009) further extends the idea of technology as not only limited to tangible tools and espouses three perspectives: fulfilment of a human need, collection of practices and processes, and aggregation of available tools. Berger et al. (2018, p. 3), deriving from Arthur (2009), build a three-layer framework of technologies as "concept", "approach", and "physical components." The concept layer embodies numerous principles that accomplish a human need and is solution agnostic. The second layer, approach, is positioned within one specific principle and can have multiple implementation approaches. The third layer, physical components, is an assemblage of specific physical or virtual purpose-built solutions to achieve the approach technological layer.

DTs as a subset of technology can be defined as "multiple principles … for the usage of digital resources … to effectively find, analyse, create, communicate, or use information in specific contexts" (Berger et al., 2018, p. 4). Berger et al. (2018) adopt the "approach" layer of technology and develop a taxonomy consisting of seven DT archetypes: platforms, connectivity, actor-based products, analytical insight generation, and augmented interaction. We build on Yoo et al. (2010)'s modular DT architecture mapping the four DT layers (service, content, network, and device) to our ontological framework as shown in Table 3.

The physical domain consisting of the device layer can be divided into physical machinery layer (e.g., computer hardware) and a logical capability layer (e.g., operating system). The network layer is divided into a physical transport layer (cables, radio spectrum, transmitters) and a logical transmission layer (network standards such as TCP/IP or peer-to-peer protocols) providing interfaces between ontological domains. The service layer deals with application functionality that directly serves users as they create, manipulate, store, and consume content. The content layer includes data (texts, sounds, images, and videos) that are stored and shared and provides metadata and directory information about the origin, ownership, copyright, encoding methods, content tags, geo-time stamps, and so on (Benkler, 2006; Farrell & Weiser, 2003).

[Table 3: DT modular architecture mapped to the ontological framework]

For this paper, AI is defined as machines or "assemblage of technological components" (Canhoto & Clear, 2020, p. 184) that perform cognitive functions associated with human minds, operate autonomously without human intervention, and learn and identify patterns to make decisions (Raisch & Krakowski, 2020; Sousa et al., 2019; von Krogh, 2018).

Adopting Berger et al. (2018) and Arthur (2009) technology layers, AI can be conceptualised at the concept level encompassing multiple principles that serve the purpose of

automating and augmenting cognitive tasks performed by humans (Raisch & Krakowski, 2020). Thus, following one specific principle, 'use of AI in digital technologies', we evaluate the ethical implications of the use of AI at the approach level. The DT archetypes, also at the approach level, provide optimal conceptual clusters of digital technologies for our purposes. The ontological framework serves as a common taxonomy between ethical implications and DT architectural layers enabling mapping to DT archetype clusters.

3 Research Methodology

3.1 Systematic Literature Review

We followed the widely used 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses' (PRISMA) (Moher et al., 2009) methodology to conduct a systematic literature review and qualitative synthesis. The objective of the review was to conduct a thematic analysis of academic literature to achieve theoretical saturation on digital ethics implications for the use of AI in DT archetypes. We extend the existing AI ethics landscape that is either focused on the grey literature synthesising policy documents (Floridi & Cowls, 2019; Jobin et al., 2019) or academic debates on AI within the context of algorithms (Mittelstadt et al., 2016). The review in this paper focuses on digital ethics implications at the DT archetype level, thus, providing a second layer of granularity.

A literature search was conducted in four phases following the PRISMA methodology. During identification, a scoping literature review was conducted to develop an understanding of the current state of the literature and test a range of keywords. A combination of two keywords, '"business ethics" AND (AI OR "artificial intelligence")' and '"digital ethics" AND (AI OR

"artificial intelligence")', were used to conduct a literature search in three databases[1], EBSCO Host, SCOPUS and ISI Web of Knowledge. The search terms were derived from the research question exploring the use of AI and corresponding digital ethics implications. In addition to the 'digital ethics' construct, for a broader search, we used 'business ethics' to encapsulate literature derived from applied ethics but within the context of organisational use of digital technologies. A total of 195 records were identified. The inclusion criteria for the search were peer-reviewed studies that discussed one or more DT archetypes and their related ethical challenges, academic publications, English language, and year of publication after 2000 to focus on the current AI research. Furthermore, through citation review and Google Scholar search[2] (first five result pages) we identified additional 42 records. The screening of the records was conducted through title and abstract review and resulted in 108 papers for the full-text article review. Out of 108 screened papers, 59 papers (shown in Appendix 1) were finally included in the qualitative synthesis. The exclusion criteria were studies that focused on the application of AI with no discussion on ethics, technical papers focussing on AI approaches and robotic applications, and papers on AI development. Figure 2 shows the PRISMA flow.

[Figure 2: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow]

Qualitative synthesis was conducted using template analysis. This enabled the development of conceptual themes, their clustering into constituent themes, and identification

---

[1] The search on EBSCO Host and ISI Web of Knowledge was conducted on September 26, 2020. The search was rerun, and Scopus was added as a new database on November 9, 2020.
[2] Conducted on November 9, 2020

across cases as global themes (King, 2004). Data analysis was conducted in three steps described below. The unit of analysis was the DT archetype.

In step one, a priori template was developed adopting the AI principles outlined in the literature (Floridi & Cowls, 2019; Jobin et al., 2019) as beneficence, non-maleficence, justice, autonomy, and explicability. In step 2, the DT archetype was identified using the modular architecture shown in Table 3 and the related digital ethical implications discussed in the literature were coded. As the ethical issues were coded, codes were grouped into organising and global themes (Attride-Stirling, 2001). The final template was produced after a few more rounds of reorganising themes and sub-themes and performing reflexivity checks. In step 3, a conceptual map was developed to summarise the results and interpret the findings.

3.2   Results

The systematic literature review included 59 papers from 43 different journals (Table 4). The comprehensive array of journals show the multi-disciplinary nature of ethical debates on AI in technology, computing, philosophy, law, nursing, medicine, management, finance, and accounting. Of the publications included in this review, over 91% were published since 2018 (Table 5) demonstrating the contemporaneous nature of the debate on ethics and AI. Figure 3 shows a wordle for all the papers included in the systematic literature review. The wordle shows data, ethics, technology, systems, researching, informs, socially, algorithms, human, and decisive as the top ten frequent stemmed words.

[Table 4: Publications included in the literature review]

[Table 5: Year of publications included in the systematic literature review]

[Figure 3: Systematic literature review papers' wordle]

All the seven DT archetypes were identified in the literature review as shown in Table 6. Analytical Insight Generation is the most prevalent archetype discussed in 76% of the papers. This

is followed by Sensor-based Data Collection at 27%, Actor-based Product at 25%, Analytical Interaction at 22%, Platform at 20%, Connectivity at 12%, and Augmented Interaction clusters at 10%.

[Table 6: DT Archetypes identified in literature review]

Table 7 and Figure 4 show the digital ethics implications inductively derived from template analysis and clustered into a priori AI principles with the identification of a sixth principle as "governance" relevant for the current context of digital technologies. The digital ethics implications are inductively clustered around four global themes that comprise the dimensions of our ontological framework developed in Section 3. The results of template analysis are discussed below.

[Table 7: Digital ethics themes identified from the qualitative synthesis]

[Figure 4: Ethical implications mapped to the ontological framework and DT architecture]

Physical Domain

The physical domain refers to Popper (1979)'s World I encompassing the device layer of DTs and shows a prevalence of ethical implications related to new technology development. This domain includes implications for dignity and well-being, safety, and sustainability under the AI principle of beneficence stipulating "do no harm" (Floridi & Cowls, 2019, p. 6).

Dignity and well-being are defined as safeguarding human dignity and promoting the well-being of the people (Floridi & Cowls, 2019; Malkiat et al., 2019). This is recognised as a key human rights principle in several policy guidelines on AI development and relates to a critical implication when balancing ethical implications of autonomy and privacy (Gregory & Halff, 2020; Malkiat et al., 2019; Vesnic-Alujevic et al., 2020).

Safety is defined as protection against physical harm from AI-enabled technologies (O'Sullivan et al., 2019; Rhim et al., 2020). This implication relates to ensuring human-machine interactions are physically safe and mitigate instances of fatalities of human actors.

Sustainability is defined as promoting the well-being of the planet and a positive view of the future (Floridi & Cowls, 2019; Neubert & Montañez, 2020). This implication is concerned with the environmental impact of new digital technologies, sustainability of the planet, and respect for public goods. It espouses a positive future outlook and the capability of AI to better serve climate and environment sustainability challenges.

Cognitive Domain

The cognitive domain referring to Popper (1979)'s World II encompasses the digital technologies service layer and the dominant ethical implications are related to machine learning and algorithms. It includes ethical implications of intelligibility and accountability related to the AI principle of explicability, autonomy related to the same AI principle, and fairness, promoting prosperity, and solidarity related to the AI principle of justice (Floridi & Cowls, 2019, pp. 7-8).

Intelligibility is defined as "the epistemological sense of how the AI works and its accuracy" (Floridi & Cowls, 2019, p. 8). Scholars associate this implication to building trust in AI outputs by ensuring that algorithms and digital/physical outputs are accurate and reliable, and machine learning is driven by causality rather than by correlations. In addition, the algorithms need to be scalable and generalisable to develop wide-ranging software applications. Furthermore, back-box AI design is extensively debated with the need for algorithms to be transparent, interpretable, and explainable to retain human agency and autonomy.

Accountability is defined as "the ethical sense of who is responsible for the way AI works" (Floridi & Cowls, 2019, p. 8). In tandem with intelligibility implication, there are extensive debates

in the literature on who is ultimately responsible for algorithmic decisions, software engineers, organisations, or machines. In a similar context, liability and culpability resulting from autonomous operations of the DTs are widely discussed.

Fairness is defined as all humans should be treated equally and the use of AI should not result in unfair discrimination against individuals, communities, or groups (Lodders & Paterson, 2020; Vidgen et al., 2020). This implication is extensively discussed in the context of AI development and algorithms to ensure the outputs are not biased even though the data used for machine learning may reflect social biases. On a positive note, scholars (Floridi, 2018; Martin, 2019b; Neubert & Montañez, 2020) discuss AI's emancipatory power to create a fair society by helping reduce societal asymmetries and racial and gender stereotypes.

Promoting prosperity is defined as enhancing the common good, being socially beneficial, and considering all stakeholders (Metcalf et al., 2019; Neubert & Montañez, 2020; Vidgen et al., 2020). This ethical implication is discussed from a utilitarian and common good argument (Vidgen et al., 2020) with the ultimate goal of AI development being beneficial for humanity (Sutton et al., 2018; Wright & Schultz, 2018).

Solidarity is defined as moral sensitivity, empathy, and appreciation for human rights (Siebecker, 2019). This ethical implication relates to fairness and prosperity focussing on the human rights principle and a call for developers to ensure empathy, social justice, and human rights for all stakeholders are considered.

Autonomy is defined as the "development of AI should not hurt the power of humans to decide" (Floridi & Cowls, 2019, p. 7). This is a key implication extensively discussed in AI development ensuring the preservation of human agency and freedom of choice to avoid humans becoming subservient to algorithmic design.

Information Domain

The information domain refers to Popper (1979)'s World III concept encompassing the content layer of digital technologies. The key ethical implications in this domain relate to privacy and security under the non-maleficence AI principle (Floridi & Cowls, 2019, p. 6).

Privacy is defined as consent and safeguards against surveillance and big data collection (Frizzo-Barker et al., 2020; Schappert & von Hauff, 2020). The primary implication revolves around access, consent, and rights to personal data in the age of big data and mass surveillance, and the use of such data in applications not disclosed during data consent. The related implication of security defined as data protection and managing information security (Stahl et al., 2016; Urquhart et al., 2019) is concerned with safeguards towards ensuring privacy and data protection.

Governance Domain

The AI principle of governance is defined as "the practice of establishing and implementing policies, procedures, and standards for the proper development, use, and management of the infosphere" (Floridi, 2018, p. 3). This domain encompasses the (in)formal rules and social moral values expressed in implications for regulatory, financial and economic, and individual and societal impact.

Regulatory impact is defined as "relevant legislation, a system of laws elaborated and enforced through social or governmental institutions to regulate the behaviour of the relevant agents in the infosphere" (Floridi, 2018, p. 3). These impacts are discussed in terms of rules and regulations to ensure human rights, intellectual property rights, data governance, and lawful surveillance.

The financial and economic impact is defined as impacts to organisations and nations as a result of AI diffusion (Grewal et al., 2020; Wang & Siau, 2019). These are discussed in both

negative and positive terms ranging from market dominance, additional revenues, and cost savings to concerns on data monetisation and antitrust.

Individual and societal impact is defined as the "transformational changes in society and individual's agency as a result of AI" (Akter et al., 2020; Schappert & von Hauff, 2020). In tandem with other implications, these impacts relate to transformational shifts in society with new technologies, labour displacement, and unemployment, and employee deskilling.

4 Discussion

The objective of this paper was to outline key digital ethics implications in the context of AI use in DTs. Through systematic literature review and qualitative synthesis, we identified 14 digital ethics implications associated with four ontological domains and mapped them to seven DT archetypes as shown in Tables 7 and 8.

Table 8 shows a conceptual mapping of 14 digital ethics implications mapped to the seven DT archetypes. Deriving from qualitative synthesis, the colour coding reflects the prominence of a given digital ethics implications for a DT archetype. Black cells reflect an ethical implication that is mentioned in more than 2/3rd of the publications that discussed a particular archetype, grey cell indicates ethical implications in 1/3rd to 2/3rd of the publications, and light grey for less than 1/3rd publications. The digital ethics implications of intelligibility, accountability, fairness, and autonomy under the cognitive domain (related to the DT service layer) and privacy under the information domain (related to the DT content layer) are the most widely discussed in our sample. Furthermore, ethical implications related to the governance domain are shown to be generally applicable for most DT archetypes. The results are no surprise given the key concern regarding AI diffusion relates to ensuring intelligibility, accountability, and fairness in algorithmic decision-making and privacy concerns related to big data applications. As well as socio-technical progress

and societal change calls for imperative debates on economic and individual impacts in policy, law and academic circles. Digital ethics implications under the physical domain (related to DT device layer) are less prominent when it comes to AI diffusion with one exception. The safety implication is worth noting for DT archetypes related to Actor-based, Analytical Interaction, and Augmented Interaction.

[Table 8: Conceptual map of digital ethics implications for DT archetypes]

Vesnic-Alujevic et al. (2020) thematic analysis of European policy documents on AI reveals nine issues, identified within our 14 digital ethical implications, categorised under individual and societal impacts of AI. Similarly, Whittlestone et al. (2019) distinguish between societal and individual values when discussing key tensions resulting from the use of AI. Jaume-Palasi (2019) draws on the work of Foucault and argues AI represents a "new form of … immaterial infrastructure" (p. 479) that creates a "second layer of norms in mathematical modules … impacting societies architectonically" (p. 483) and hence needs to be analysed at the societal level even though incorporating individual values on the surface.

DT archetypes through their inherent architectural requirements and design motives might need to prioritise one domain or ethical implication over another. Such prioritisation represents a "true dilemma", a conflict whereby ethical implication(s) are inherently conflicting, or a "dilemma in practice", a conflict arising from resource constraints, or a "false dilemma" involving failures in recognising true impacts (Whittlestone et al., 2019, p. 42). Literature suggests an enhanced form of cost-benefit analysis involving all stakeholder voices and public deliberations as a potential solution for resolution of "true" and "in practice" dilemmas (Ibid.). Considering AI as an "immaterial infrastructure" (Jaume-Palasi, 2019, p. 479) and incorporating cost-benefit analysis at the societal level, the fundamental reflection is to analyse identified digital ethics implications and

their associated tensions in terms of the overall societal impact. We define societal impact in our context modifying the definition developed by Bornmann (2013, p. 217) as "the assessment of social, cultural, environmental, and economics returns (impact and effects) from [AI enabled digital products and services and associated human-machine interactions]."

Thus, we state our first proposition as:

Proposition 1: The use of AI in DTs involves resolving ethical implications in the physical [P1A], cognitive [P1B], information [P1C], and governance [P1D] domains that affect the overall societal impact.

The economic value realisation is the prime driver for the use of AI in DTs within organisations. Chesbrough (2010) argues firms need to innovate their business models to monetise new technologies. Thus, from an organisational perspective, the use of AI in DTs is primarily driven by economic values of revenue generation and cost savings. The organisational culture, leadership, and its business model operating within the confines of the regulatory boundaries determine value generation from the use of AI in DTs.

ISO:26000 (2010, p. 3) defines social responsibility as "the responsibility of an organization for the impacts of its decisions and activities on society and the environment through transparent and ethical behavior." Stakeholder theory stipulates organisations need to be cognizant of the diverse and contradictory needs of all their stakeholder groups (Freeman, 1984). The new generation of stakeholders is increasingly conscious of social responsibility. Organisations that adopt socially responsible practices will have greater economic success with the use of AI (Du & Xie, 2020) and improve their competitive position.

The design characteristics of new technology are strongly influenced by the dominant market players in addition to the culture and technological capabilities of the implementing

organisation. Crawford (2021, p. 8) states that AI in this sense is "a registry of power." This institutional environment affects how organisations situate their social responsibility efforts and organisational cultural norms for value realisation from the use of AI in DTs. Thus, we state our second proposition as:

Proposition 2: The organisational impact mediates the effect of physical [P3A], cognitive [P3B], information [P3C], and governance [P3D] domains on the overall societal impact.

The prevalence of the Analytical Insights Generation archetype in our review (Table 5) demonstrates the imperative discussions on algorithms and machine learning as the most critical within the context of digital ethics and AI applications. This is followed by the increasing use of AI in Sensor-based and Actor-based technologies in healthcare, smart homes, autonomous vehicles, etc. Both Platform and Connectivity as infrastructure technologies invisible to end users are not actively discussed even though AI development in these clusters is quite active such as blockchains, cloud computing, financial platforms, etc. The clusters of Analytical Interaction and Augmented Interaction are a growing field of digital technologies made possible by advancements in AI, however, these are also not yet widely discussed archetypes in literature within the context of digital ethics. Using the conceptual mapping in Table 8, we highlight eight key results specific to each DT archetype.

Platform-oriented DT archetypes are characterised as infrastructure technologies acting as hubs transmitting data from a single source with a one-to-many design (Berger et al., 2018). Given data is a pivotal component of this archetype, our conceptual map highlights content being the key architecture layer when using AI in this DT archetype. Connectivity DT archetypes also characterised as infrastructure technologies add another dimension to data transmission with

efficient throughput objectives (Berger et al., 2018). Thus, our conceptual map shows the use of AI in this archetype not only focuses on the content layer but also the service layer.

Actor-based Product DT archetypes are characterised as application technologies with the primary function related to data execution from digital input to a physical output (Berger et al., 2018). Thus, our conceptual map shows the use of AI in this archetype focuses on both the device layer, with the ability to impact the physical environment, and the service layer, as the cognitive engine.

Sensor-based Data Collection DT archetypes are characterised as application technologies with the primary function of data collection from physical input to digital output (Berger et al., 2018). Thus, our conceptual map shows the use of AI in this archetype is primarily focused on data and the content layer.

Analytical Insight Generation DT archetypes are characterised as application technologies with logic and decision-making functions (Berger et al., 2018). Our conceptual map showcases the use of AI in this archetype, the highest percentage in our sample, leverages both big data and advanced machine learning capabilities to replicate human cognitive functions focussing on content and service layers.

Analytical Interaction DT archetypes are characterised as application technologies associated with the transmission of digital input, analytical capabilities, and physical output (Berger et al., 2018). Our conceptual map shows the use of AI in this archetype focuses on the content layer with data transmission function and is often associated with replacement of service jobs leading to individual and societal impacts.

Augmented Interaction DT archetypes are associated with data collection allowing for a physical input to a digital output and "embodies interaction capabilities without deeper analytical

capabilities" (Berger et al., 2018, p. 11). Our conceptual map shows the use of AI in this archetype focuses on the content layer with data primacy functions, service layer with physical interaction capabilities, and its application leading to macro shifts in the industry.

Thus, we argue the characteristics of DT archetypes moderate the effect of ontological domains on societal impact and state our third proposition as:

Proposition 3: DT archetypes moderate the effect of physical [P2A], cognitive [P2B], information [P2C], and governance [P2D] domains on the overall societal impact.

The conceptual model derived from these propositions is shown in Figure 5. The constructs of physical, cognitive, information, and governance domains are suggested as reflective that may be measured through the corresponding ethical implications. The organisational impact can be measured in terms of measures associated with performance such as profits, throughput, customer satisfaction, etc. or a hybrid scale consisting of performance measures and social responsibility measures.

[Figure 5: Conceptual Model for the digital ethical assessment of the use of AI in DTs]

4.1 Theoretical Contributions

This paper extends the current systematic literature review in AI (Borges et al., 2020; Duan et al., 2019; Dwivedi et al., 2021b; Kankanhalli et al., 2019; Nishant et al., 2020; Sarker et al., 2019) and AI ethics (Floridi & Cowls, 2019; Jobin et al., 2019; Mittelstadt et al., 2016) with five theoretical contributions. First, the paper develops a new typology embedded in an ontological framework to conceptualise the ethical impact of AI on seven DTs archetypes identified in the literature. Second, the ontological framework identifies an additional AI principle of governance missing from the current AI principles debates. The governance domain is critical within the context of DTs and exploration of the interrelations between regulations and soft and hard ethics (Floridi, 2018). Third,

deriving from an applied ethics perspective, the paper identifies eight ethical research domains in DTs and positions digital ethics as the focal point when conceptualised through our ontological framework. Fourth, through systematic literature review and qualitative synthesis, we outline 14 digital ethics implications associated with the four ontological domains and mapped them to the seven DT archetypes (conceptual map). Fifth, the conceptual model and twelve propositions deducted from our analysis can be tested and used as an empirical instrument for analysing the ethical use of AI in different contexts.

4.2 Managerial Contributions

The intent of qualitative synthesis was theoretical saturation and hence Table 8 provides the breath of ethical implications that practitioners need to consider when conducting AI ethical impact analysis on a specific DT archetype. The colour coding provides a subjective guideline on the priorities based on our sample of reviewed studies. The associated conceptual map (Table 8) provides an additional instrument that can be used for prioritising tensions related to ethical conflicts. For example, a developer interested in applications related to Sensor-based Data Collection may start their analysis by focussing on the ethical implication pertinent to that DT archetype. The highest implication (cells in black followed by grey) for this archetype is related to privacy followed by intelligibility and fairness. These key implications should be explored and balanced with the impacts of technology under the governance domain related to regulatory and societal implications. An overall societal impact measure can assist with deciding on balancing negative outcomes related to compromising on the physical domain that may conflict with satisfying higher priority ethical implications under cognitive, governance, and information domain.

4.3 Limitations and Future Research

This study does come with some limitations which can pave the path for future research avenues. First, the intent of this review was theoretical saturation and not an extant literature review. Hence, any conclusions concerning the count and priority should be considered cautiously and within the context of the DT archetypes being developed. Second, by limiting analysis at the "approach" level of technology we intended to capture the majority of DT clusters at an abstract level. However, with the exceedingly rapid progress in the technological sector, new DT clusters might arise and will need to be added to the model. Furthermore, the archetype levels do not guarantee mutually exclusive taxonomy and some DTs might fall into two archetypes. For example, the use of AI for natural language processing falls under both Sensor-based Data Collection (a wearable that can translate languages) and Analytical Insights Generation (for conducting sentiment and emotional analysis). Third, ethical impact analysis is a complex area that requires insights from several disciplines such as technology, medical ethics, policy, law, philosophy, etc. Not least, tensions between ethical implications are due to arise (Whittlestone et al., 2019) that will require extensive policy and academic debates. The conceptual model developed in this paper can be developed as an instrument through in-depth interviews and quantitative testing to conduct ethical AI analysis in different contexts.

5   Conclusions

Technological progress in the use of AI in DTs has intensified since the start of this century. Applications driven by AI are now embedded in all areas of human existence and the concept of machine general intelligence being able to pass the Turing Test seems inevitable in the next few decades if not years. At the same time, several mishaps in AI propagating biases, stereotypes, and inexplicable autonomous decisions have highlighted an urgent need for the global community to

direct the ethical development of AI in current and future technologies. Industry and governments have converged on key AI principles but a lack of clarity and divergence on the implementation remains. Adopting an applied ethics perspective, we identified digital ethics as the focal point of eight ethical research domains in DTs. To our knowledge, this is the first paper that develops an ontological framework consisting of four domains (physical, cognitive, information, and governance) to conceptualise the ethical impact of AI on the DT archetype level and introduces governance as a new AI principle. Through a systematic literature review, we identified 14 digital ethics implications associated with the four ontological domains and mapped them to seven DT archetypes. The resulting conceptual model is geared towards academics and practitioners to be able to conduct AI ethical analysis on DTs and test the twelve corresponding propositions to further validate the instrument.

Table 1: Ontological frameworks for information science

Ontological Framework	Popper (1979)	Ogden and Richards (1923); Project and Peirce (1998)	Denning and Rosenbloom (2009); Rosenbloom (2012)
Physical domain	World I	Referent/object	Physical
Cognitive domain	World II	Reference/interpretant	Social
Information domain	World III	Symbol/sign	Life

Table 2: Research domains in applied ethics in digital technologies

Ethical research domain	Definition	Ontological domain
Technoethics	ethics involved in the development of new technology and how technology alters the power of individuals (Jonas, 1985)	Physical
Computer ethics	analysis of nature and social impact of computer technology and corresponding formulation and justification of policies for the ethical use of such technology (Moor, 1985)	Physical Information
Information ethics	relationship between the creation, organisation, dissemination, and use of information, and the ethical standards and moral codes governing human conduct in society ((Joan, 2010)	Physical Information
Roboethics	study, understand and regulate the ethical implications and consequences of robotics technology; focus on moral design, development, and use of robots for the benefit of humanity (Tzafestas, 2018)	Physical Cognitive
Machine ethics	is concerned with how machines behave toward human users and other machines (Anderson & Anderson, 2006)	Physical Cognitive
Computational ethics	is an agent-based simulation mechanism that takes a computational perspective to ethics theory (Ruvinsky, 2007)	Cognitive Informational
ICT Ethics	"ICT affect human attitudes, values and behavior, and set conditions for human identity, existence, and change" (Karlsson, 2003, p. 79)	Physical Information
Artificial Intelligence – Machine Learning ethics	concerns with responsibility, transparency, auditability, incorruptibility, predictability; morality of machines (Bostrom & Yudkowsky, 2014)	Cognitive

Table 3: DT modular architecture mapped to the ontological framework

Ontological domain	DT Modular Architecture (Berger et al., 2018; Yoo et al., 2010)	DT Archetypes Properties (Berger et al., 2018)
Physical (P)	Device layer: physical machinery and logical capability	Role of technology: infrastructure or application
Cognitive (C)	Service layer: applications and algorithms	Human involvement: active or passive
Information (I)	Content layer: data	Input and outputs: digital or physical
Governance (G)	-	Governance of device (G&P): centralised or decentralised Governance of content (G&C): open or closed Governance of information (G&I): public or private
Interfaces	Network layer	Multiplicity (number and order of network nodes): one to many, many to many The direction of information flow: bi-directional or uni-directional Data treatment: collection, analysis, or transmission

Table 4: Publications included in the literature review

Publication	Count	Percentage
International Journal of Information Management	8	14%
Journal of Business Ethics	4	7%
Business Horizons	3	5%
AI & Society	2	3%
International Review of Law, Computers & Technology	2	3%
Journal of Emerging Technologies in Accounting	2	3%
Telecommunications Policy	2	3%
ACM Computing Surveys	1	2%
Alternative Law Journal	1	2%
Berkeley Technology Law Journal	1	2%
Canadian Association of Radiologists Journal	1	2%
Computers in Human Behavior	1	2%
European Journal of Operational Research	1	2%
Information and Organization	1	2%
Information Technology & People	1	2%
Intelligent Systems in Accounting, Finance & Management	1	2%
International Journal of Medical Informatics	1	2%
Journal of Behavioral and Experimental Finance	1	2%
Journal of Business Research	1	2%
Journal of Corporation Law	1	2%
Journal of International Technology & Information Management	1	2%
Philosophy and Technology	1	2%
Journal of Database Management (JDM)	1	2%
Journal of Economic Psychology	1	2%
Journal of Interactive Marketing	1	2%
Journal of Retailing and Consumer Services	1	2%
Journal of Revenue and Pricing Management	1	2%
Journal of Service Management	1	2%
Minds and Machines	1	2%
MIS Quarterly Executive	1	2%
NPQ: New Perspectives Quarterly	1	2%
Public Relations Review	1	2%
Records Management Journal	1	2%
Research Technology Management	1	2%
Science and Engineering Ethics	1	2%
Seminars in Oncology Nursing	1	2%

Publication	Count	Percentage
Social Responsibility Journal	1	2%
Social Science & Medicine	1	2%
Technological Forecasting & Social Change	1	2%
Telematics and Informatics	1	2%
The International Journal of Medical Robotics and Computer Assisted Surgery	1	2%
University of Chicago Law Review	1	2%
ZfWU - Journal for Business, Economics & Ethics	1	2%

Table 5: Year of publications included in the systematic literature review

Year	Count	Percentage
2008	1	1.7%
2014	1	1.7%
2015	1	1.7%
2017	2	3.4%
2018	5	8.5%
2019	18	30.5%
2020	30	50.8%
2021	1	1.7%

Table 6: DT Archetypes identified in literature review

DT Archetype	% of papers (59) mentioning the DT archetype	Codes	Sources
Platform	20%	Internet search, full-service wealth management with e-advisory, clinical trial management system, electronic health, telecare, health and fitness platform, social media platform	(Barlow, 2020; Breidbach & Maglio, 2020; Cath et al., 2018; Dhagarra et al., 2020; Du & Xie, 2020; Flick et al., 2020; Gerlick & Liozu, 2020; Giddens, 2018; Neubert & Montañez, 2020; Roberts et al., 2021; Savirimuthu, 2017; Shanmuganathan, 2020)
Connectivity	12%	Blockchain, electronic records management, electronic document sharing, cloud computing	(Barlow, 2020; Dhagarra et al., 2020; Frizzo-Barker et al., 2020; Gerlick & Liozu, 2020; Giddens, 2018; Yong et al., 2019)
Actor-based Product	25%	autonomous vehicles, smart house, robotics, autonomous weapon system, automated stock trading, adaptive architecture, autonomous robotic surgery	(Breidbach & Maglio, 2020; Du & Xie, 2020; Flick et al., 2020; Guidotti et al., 2018; Headrick, 2014; Johnson, 2015; Neubert & Montañez, 2020; Nicodemo & Cardoso, 2019; O'Sullivan et al., 2019; Rhim et al., 2020; Roberts et al., 2021; Robles Carrillo, 2020; Urquhart et al., 2019; Wang & Siau, 2019; Wright & Schultz, 2018)
Sensor-based Data Collection	27%	Facial recognition, smart homes, wearable devices, health sensing, sensors, biometric surveillance, mobile phone tracking, natural language processing	(Breidbach & Maglio, 2020; Dhagarra et al., 2020; Du & Xie, 2020; Ebert, 2019; Flick et al., 2020; Gerlick & Liozu, 2020; González-Rodríguez et al., 2020; Gregory & Halff, 2020; Ioannou et al., 2020; Lodders & Paterson, 2020; Jessica Morley et al., 2020; Nicodemo & Cardoso, 2019; O'Leary, 2019; Roberts et al., 2021; Savirimuthu, 2017; Wright & Schultz, 2018)

DT Archetype	% of papers (59) mentioning the DT archetype	Codes	Sources
Analytical Insight Generation	76%	Machine learning, big data analytics, data mining, deep learning, neural networks, algorithms, people analytics, bots, decision systems, recommender systems, natural language processing	(Barlow, 2020; Breidbach & Maglio, 2020; Burk, 2019; Cho et al., 2020; Dhagarra et al., 2020; Du & Xie, 2020; Duan et al., 2019; Ebert, 2019; Flick et al., 2020; Gal et al., 2020; Geis et al., 2019; Gerlick & Liozu, 2020; Giddens, 2018; González-Rodríguez et al., 2020; Gregory & Halff, 2020; Guidotti et al., 2018; Ioannou et al., 2020; Lee & Shin, 2020; Leicht-Deobald et al., 2019; Lodders & Paterson, 2020; Malkiat et al., 2019; Martin, 2019a, 2019b; Milano et al., 2020; Jessica Morley et al., 2020; Mraović, 2008; Mulligan & Bamberger, 2019; Munoko et al., 2020; Neubert & Montañez, 2020; Nicodemo & Cardoso, 2019; O'Leary, 2019; Roberts et al., 2021; Robles Carrillo, 2020; Savirimuthu, 2017; Schniter et al., 2020; Shanmuganathan, 2020; Siebecker, 2019; Sutton et al., 2018; Vesnic-Alujevic et al., 2020; Vidgen et al., 2020; Wang & Siau, 2019; Watson, 2019; Wright & Schultz, 2018)
Analytical Interaction	22%	human enhancement technology, anthropomorphism, interactive avatars	(Cath et al., 2018; Chen et al., 2020; Du & Xie, 2020; Grewal et al., 2020; Manfreda et al., 2021; Jessica Morley et al., 2020; Nicodemo & Cardoso, 2019; Roberts et al., 2021; Schniter et al., 2020; Sipior, 2020; Subramanian, 2017; Sutton et al., 2018; Wang & Siau, 2019; Wiesenberg & Tench, 2020; Wright & Schultz, 2018; Xie et al., 2020)
Augmented Interaction	10%	human augmentation technology, adaptive architecture, robotics	(Dwivedi et al., 2021b; Grewal et al., 2020; Nicodemo & Cardoso, 2019; Roberts et al., 2021; Urquhart et al., 2019; Wang & Siau, 2019)

Table 7: Digital ethics themes identified from the qualitative synthesis

Ontological Domain	AI Principle	Ethical Implication	Codes	Sources
Cognitive	Explicability "the knowledge of how AI works and who to hold responsible for its outcomes" (Floridi & Cowls, 2019, p. 8)	C1 - Intelligibility	right to explanation, transparency, opacity, black-box design, interpretability, explainability, comprehensibility, traceability, intelligibility, accuracy, efficiency, quality, reliability, minimise error, reduce risk, detecting causality than correlations, trust in an algorithm, faith, fidelity, generality, scalability,	(Barlow, 2020; Breidbach & Maglio, 2020; Chen et al., 2020; Cho et al., 2020; Du & Xie, 2020; Duan et al., 2019; Dwivedi et al., 2021b; Flick et al., 2020; Frizzo-Barker et al., 2020; Gal et al., 2020; Geis et al., 2019; Giddens, 2018; Gregory & Halff, 2020; Grewal et al., 2020; Guidotti et al., 2018; Headrick, 2014; Lee & Shin, 2020; Leicht-Deobald et al., 2019; Lodders & Paterson, 2020; Manfreda et al., 2021; Martin, 2019a, 2019b; Milano et al., 2020; Jessica Morley et al., 2020; Mulligan & Bamberger, 2019; Munoko et al., 2020; Neubert & Montañez, 2020; Roberts et al., 2021; Robles Carrillo, 2020; Shanmuganathan, 2020; Sipior, 2020; Vesnic-Alujevic et al., 2020; Vidgen et al., 2020; Wang & Siau, 2019; Watson, 2019; Wiesenberg & Tench, 2020; Yong et al., 2019)
	Explicability	C2 - Accountability	accountability, responsibility, liability, ownership of data and decisions, culpability	(Barlow, 2020; Breidbach & Maglio, 2020; Cath et al., 2018; Chen et al., 2020; Du & Xie, 2020; Duan et al., 2019; Frizzo-Barker et al., 2020; Geis et al., 2019; Guidotti et al., 2018; Headrick, 2014; Johnson, 2015; Leicht-Deobald et al., 2019; Lodders & Paterson, 2020; Malkiat et al., 2019; Manfreda et al., 2021; Martin, 2019a, 2019b; Mulligan & Bamberger, 2019; Munoko et al., 2020; Neubert & Montañez, 2020; O'Leary, 2019; O'Sullivan et al., 2019; Roberts et al.,

Ontological Domain	AI Principle	Ethical Implication	Codes	Sources
				2021; Robles Carrillo, 2020; Shanmuganathan, 2020; Subramanian, 2017; Urquhart et al., 2019; Vesnic-Alujevic et al., 2020; Watson, 2019; Wiesenberg & Tench, 2020; Yong et al., 2019)
	Justice the quality of being fair and eliminating discrimination ensuring equal access to the benefits of AI (Floridi & Cowls, 2019)	C3 - Fairness	avoiding bias, fairness, justice, accessibility, discrimination, human rights, racial and gender stereotypes, information asymmetries, equality, freedom and justice, basic rights, equality, fair use, unfair outcomes	(Breidbach & Maglio, 2020; Burk, 2019; Cho et al., 2020; Du & Xie, 2020; Dwivedi et al., 2021b; Flick et al., 2020; Frizzo-Barker et al., 2020; Geis et al., 2019; Gerlick & Liozu, 2020; Giddens, 2018; Grewal et al., 2020; Guidotti et al., 2018; Lee & Shin, 2020; Leicht-Deobald et al., 2019; Lodders & Paterson, 2020; Martin, 2019a, 2019b; Milano et al., 2020; Jessica Morley et al., 2020; Munoko et al., 2020; Neubert & Montañez, 2020; Roberts et al., 2021; Robles Carrillo, 2020; Siebecker, 2019; Sipior, 2020; Vesnic-Alujevic et al., 2020; Vidgen et al., 2020; Wang & Siau, 2019; Yong et al., 2019)
	Justice	C4 - Promoting prosperity	Socially beneficial, prudence, human values principle, common good, augment human capabilities than replacing them, the benefit of humanity, attention to context and culture	(Malkiat et al., 2019; Mulligan & Bamberger, 2019; Munoko et al., 2020; Neubert & Montañez, 2020; Rhim et al., 2020; Roberts et al., 2021; Sutton et al., 2018; Vidgen et al., 2020; Wright & Schultz, 2018) (Cath et al., 2018; Manfreda et al., 2021)
	Justice	C5 - Solidarity	solidarity, empathy, social inequality issues, social justice	(Cath et al., 2018; Gerlick & Liozu, 2020; Grewal et al., 2020; Leicht-Deobald et al., 2019; Mraović, 2008; Munoko et al., 2020; O'Leary, 2019; Savirimuthu,

Ontological Domain	AI Principle	Ethical Implication	Codes	Sources
		moral sensitivity, empathy, and appreciation for human rights		2017; Siebecker, 2019; Sutton et al., 2018; Vidgen et al., 2020; Yong et al., 2019)
	Autonomy "development of AI should not hurt the power of humans to decide" (Floridi & Cowls, 2019, p. 7)	C6 - Autonomy	autonomy, choice, human free agency, freedom of choice, nudging, power of user	(Breidbach & Maglio, 2020; Cath et al., 2018; Du & Xie, 2020; Flick et al., 2020; Gal et al., 2020; Geis et al., 2019; Giddens, 2018; Gregory & Halff, 2020; Leicht-Deobald et al., 2019; Lodders & Paterson, 2020; Martin, 2019a, 2019b; Milano et al., 2020; Mraović, 2008; Munoko et al., 2020; O'Leary, 2019; Rhim et al., 2020; Subramanian, 2017; Sutton et al., 2018; Vesnic-Alujevic et al., 2020; Vidgen et al., 2020)
Physical	Beneficence AI should "do only good" and promote the well-being of humanity above all else (Floridi & Cowls, 2019, p. 6)	P1 - Dignity and well being	temperance, responsible leadership, stakeholder rights, human dignity, how potential users are treated, loss of agency, compassion, genuine concern, datafication of society	(Du & Xie, 2020; Dwivedi et al., 2021b; Flick et al., 2020; Geis et al., 2019; Gregory & Halff, 2020; Malkiat et al., 2019; Martin, 2019b; Munoko et al., 2020; Neubert & Montañez, 2020; Rhim et al., 2020; Vesnic-Alujevic et al., 2020; Vidgen et al., 2020; Wiesenberg & Tench, 2020)
	Beneficence	P2 – Safety (O'Sullivan et	safety, mitigate harm, reduce fatality, damage	(Du & Xie, 2020; Guidotti et al., 2018; Johnson, 2015; Manfreda et al., 2021; Munoko et al., 2020; Neubert &

Ontological Domain	AI Principle	Ethical Implication	Codes	Sources
		al., 2019; Rhim et al., 2020)		Montañez, 2020; O'Sullivan et al., 2019; Rhim et al., 2020; Roberts et al., 2021; Robles Carrillo, 2020; Urquhart et al., 2019; Wang & Siau, 2019; Wright & Schultz, 2018)
	Beneficence	P3 - Sustainability	a positive view of the future, respect for the public good, utilitarian, energy use, environmental effects, sustainability of the planet	(Dwivedi et al., 2021b; Frizzo-Barker et al., 2020; Neubert & Montañez, 2020; Roberts et al., 2021; Robles Carrillo, 2020; Vesnic-Alujevic et al., 2020; Vidgen et al., 2020)
Information	Non-maleficence "AI should do no harm and avoid misuse" (Floridi & Cowls, 2019, p. 6)	I1 - Privacy	privacy, access to personal data, surveillance, digital rights, trust, consent, datafication	(Chen et al., 2020; Dhagarra et al., 2020; Du & Xie, 2020; Dwivedi et al., 2021b; Flick et al., 2020; Frizzo-Barker et al., 2020; Gal et al., 2020; Geis et al., 2019; Gerlick & Liozu, 2020; Giddens, 2018; González-Rodríguez et al., 2020; Gregory & Halff, 2020; Grewal et al., 2020; Guidotti et al., 2018; Ioannou et al., 2020; Lee & Shin, 2020; Lodders & Paterson, 2020; Malkiat et al., 2019; Manfreda et al., 2021; Milano et al., 2020; Mraović, 2008; Munoko et al., 2020; Roberts et al., 2021; Robles Carrillo, 2020; Savirimuthu, 2017; Schniter et al., 2020; Shanmuganathan, 2020; Sipior, 2020; Subramanian, 2017; Sutton et al., 2018; Urquhart et al., 2019; Vesnic-Alujevic et al., 2020; Wang & Siau, 2019; Wiesenberg & Tench, 2020; Xie et al., 2020; Yong et al., 2019)
	Non-maleficence	I2 - Security	security, data protection, confidentiality	(Du & Xie, 2020; Dwivedi et al., 2021b; Geis et al., 2019; González-Rodríguez et al., 2020; Grewal et al., 2020; Guidotti et al., 2018; Manfreda et al., 2021;

Ontological Domain	AI Principle	Ethical Implication	Codes	Sources
				Munoko et al., 2020; Neubert & Montañez, 2020; Robles Carrillo, 2020; Shanmuganathan, 2020; Sipior, 2020; Subramanian, 2017; Urquhart et al., 2019; Wang & Siau, 2019; Yong et al., 2019)
Governance	Governance "is the practice of establishing and implementing policies, procedures and standards for the proper development, use and management of the infosphere." (Floridi, 2018, p. 3)	G1 - Regulatory impact	avoid deception and coercion, policies to reduce social injustice, human rights and victim access to an effective remedy, intellectual property, data ownership, occupational rights, surveillance, consent	(Cath et al., 2018; Chen et al., 2020; Du & Xie, 2020; Duan et al., 2019; Dwivedi et al., 2021b; Gerlick & Liozu, 2020; Headrick, 2014; Lodders & Paterson, 2020; Malkiat et al., 2019; Manfreda et al., 2021; O'Leary, 2019; O'Sullivan et al., 2019; Roberts et al., 2021; Robles Carrillo, 2020; Savirimuthu, 2017; Sipior, 2020; Subramanian, 2017; Urquhart et al., 2019; Vesnic-Alujevic et al., 2020; Wang & Siau, 2019; Wright & Schultz, 2018; Yong et al., 2019)
	Governance	G2 - Financial and economic impact	positive benefits in the marketplace, reduce cost, data monetisation, additional revenues, antitrust factors, corporate digital responsibility, national and international economic impacts	(Breidbach & Maglio, 2020; Dwivedi et al., 2021b; Gerlick & Liozu, 2020; Giddens, 2018; Grewal et al., 2020; Neubert & Montañez, 2020; Roberts et al., 2021; Vidgen et al., 2020; Wang & Siau, 2019; Wiesenberg & Tench, 2020; Wright & Schultz, 2018; Yong et al., 2019)
	Governance	G3 - Individual and societal impact	shifts in society with technological advancement, change in cultural and	(Breidbach & Maglio, 2020; Cath et al., 2018; Chen et al., 2020; Du & Xie, 2020; Duan et al., 2019; Dwivedi et al., 2021b; Flick et al., 2020; Gal et al., 2020; Geis

Ontological Domain	AI Principle	Ethical Implication	Codes	Sources
			personal values, moral consequences, unemployment, retraining of displaced workers, deskilling, basic existential principles of humanity and society, social effects, the wealth gap, digital gap, job displacement and replacement, isolation, deprofessionalisation	et al., 2019; Lee & Shin, 2020; Leicht-Deobald et al., 2019; Lodders & Paterson, 2020; Martin, 2019a, 2019b; Milano et al., 2020; Jessica Morley et al., 2020; Munoko et al., 2020; Neubert & Montañez, 2020; O'Leary, 2019; Robles Carrillo, 2020; Subramanian, 2017; Sutton et al., 2018; Vesnic-Alujevic et al., 2020; Vidgen et al., 2020; Wang & Siau, 2019; Wiesenberg & Tench, 2020; Wright & Schultz, 2018)

Table 8: Conceptual map of digital ethics implications for DT archetypes

DT Archetypes	Physical domain			Cognitive domain						Information domain		Governance domain		
	Device layer			Service layer						Content layer		Governance layer		
	P1	P2	P3	C1	C2	C3	C4	C5	C6	I1	I2	G1	G2	G3
Platform	17%	17%	8%	58%	50%	50%	17%	25%	42%	67%	17%	33%	33%	33%
Connectivity	14%	14%	14%	86%	71%	86%	0%	29%	29%	86%	29%	29%	43%	14%
Actor-based Product	27%	80%	13%	60%	67%	47%	33%	0%	33%	53%	40%	60%	27%	40%
Sensor-based Data Collection	25%	19%	13%	56%	31%	56%	19%	19%	31%	81%	25%	56%	31%	44%
Analytical Insight Generation	24%	16%	11%	69%	49%	53%	22%	20%	40%	67%	20%	40%	18%	51%
Analytical Interaction	23%	38%	15%	62%	38%	54%	31%	15%	38%	85%	38%	46%	38%	69%
Augmented Interaction	17%	50%	33%	67%	50%	67%	17%	33%	17%	83%	67%	83%	67%	33%

Key:

■ the ethical implication was cited in over 2/3rd publications that mentioned a particular archetype

▓ the ethical implication was cited in between 1/3rd and 2/3rd publications that mentioned a particular archetype

░ the ethical implication was cited in less than 1/3rd of publications that mentioned a particular archetype

Code	Ethical Implications	Code	Ethical Implications
P1	Dignity and well-being	C5	Solidarity
P2	Safety	C6	Autonomy
P3	Sustainability	I1	Privacy
C1	Intelligibility	I2	Security
C2	Accountability	G1	Regulatory impact
C3	Fairness	G2	Financial and economic impact
C4	Promoting prosperity	G3	Individual and societal impact

Figure 1: Digital ethics research domain

Physical domain

Technoethics

Roboethics
Machine
ethics

Computer
ethics
ICT ethics

Digital
ethics

**Governance
domain**

AI ethics (machine
learning ethics)

Computational
ethics

Information ethics
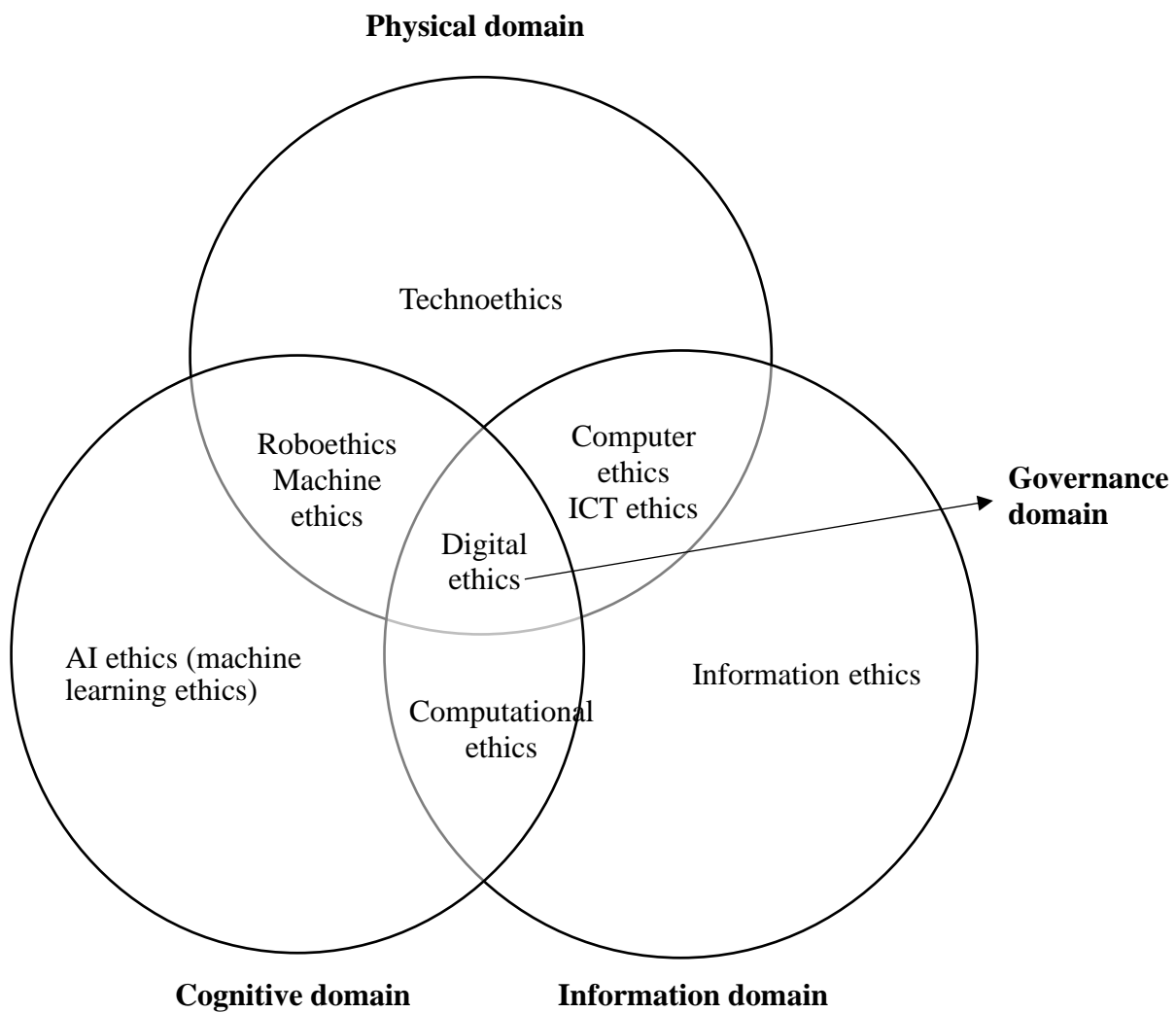
Cognitive domain   **Information domain**

Figure 2: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow
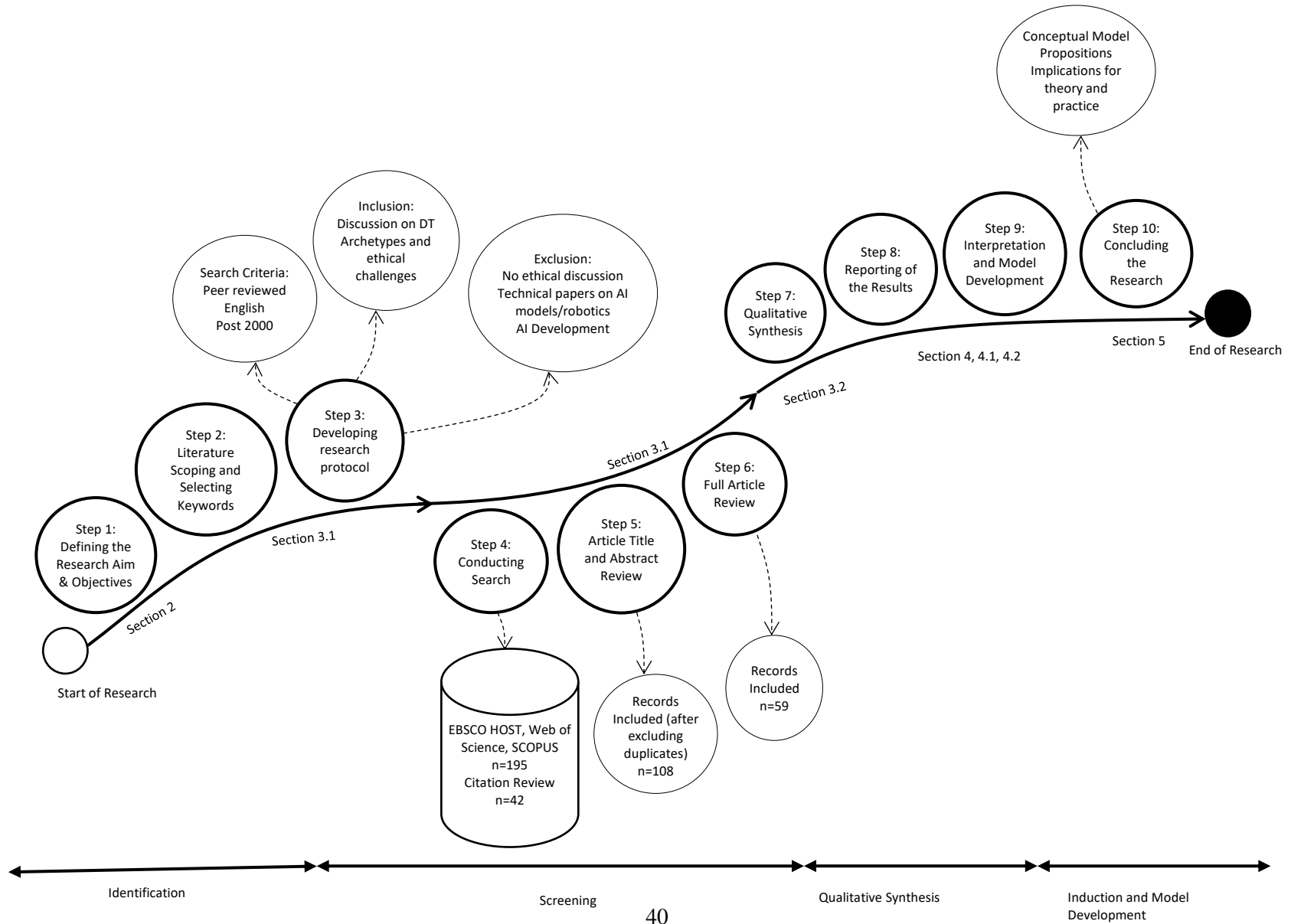
Figure 3: Systematic literature review papers' wordle

Figure 4: Ethical implications mapped to the ontological framework and DT architecture
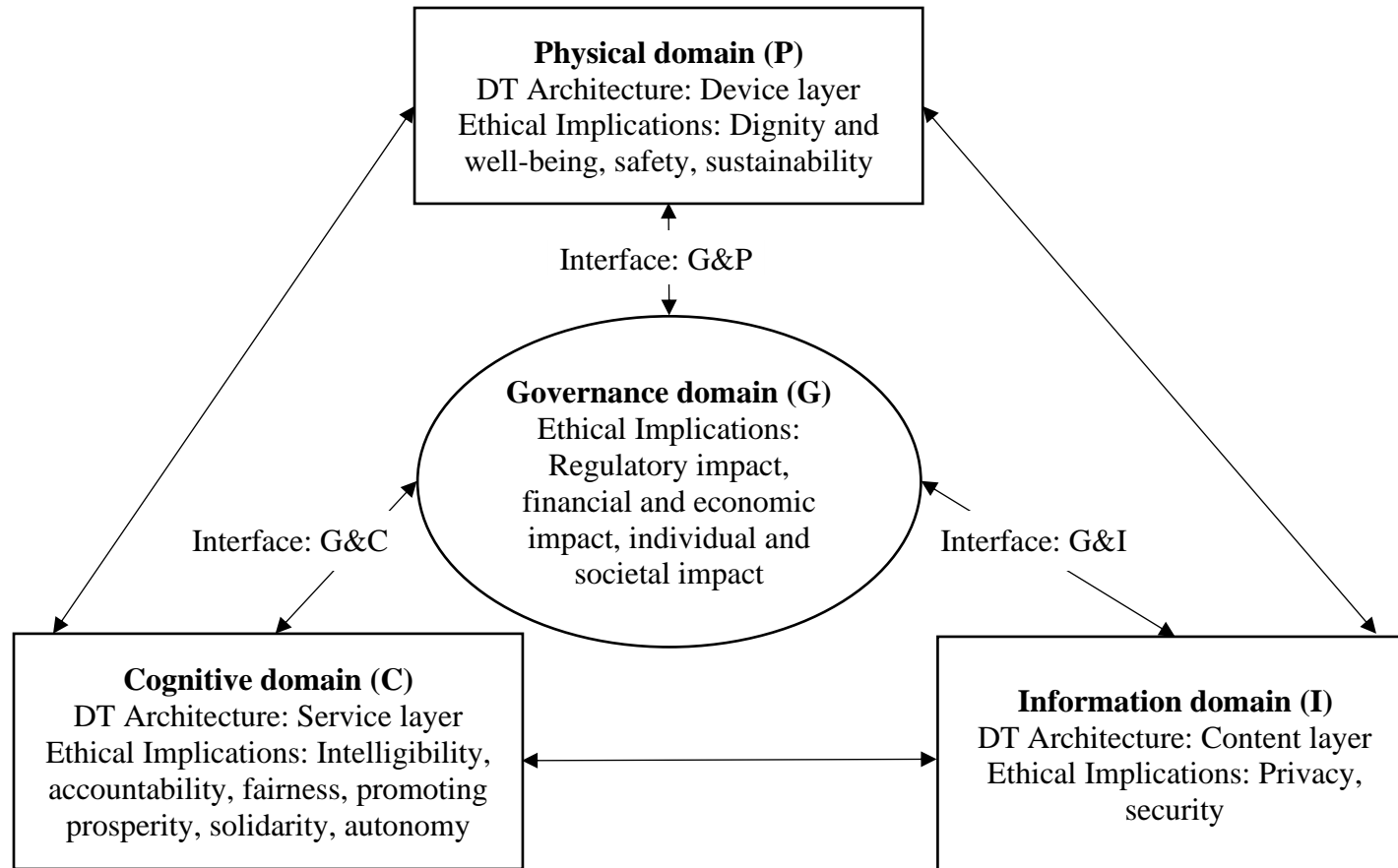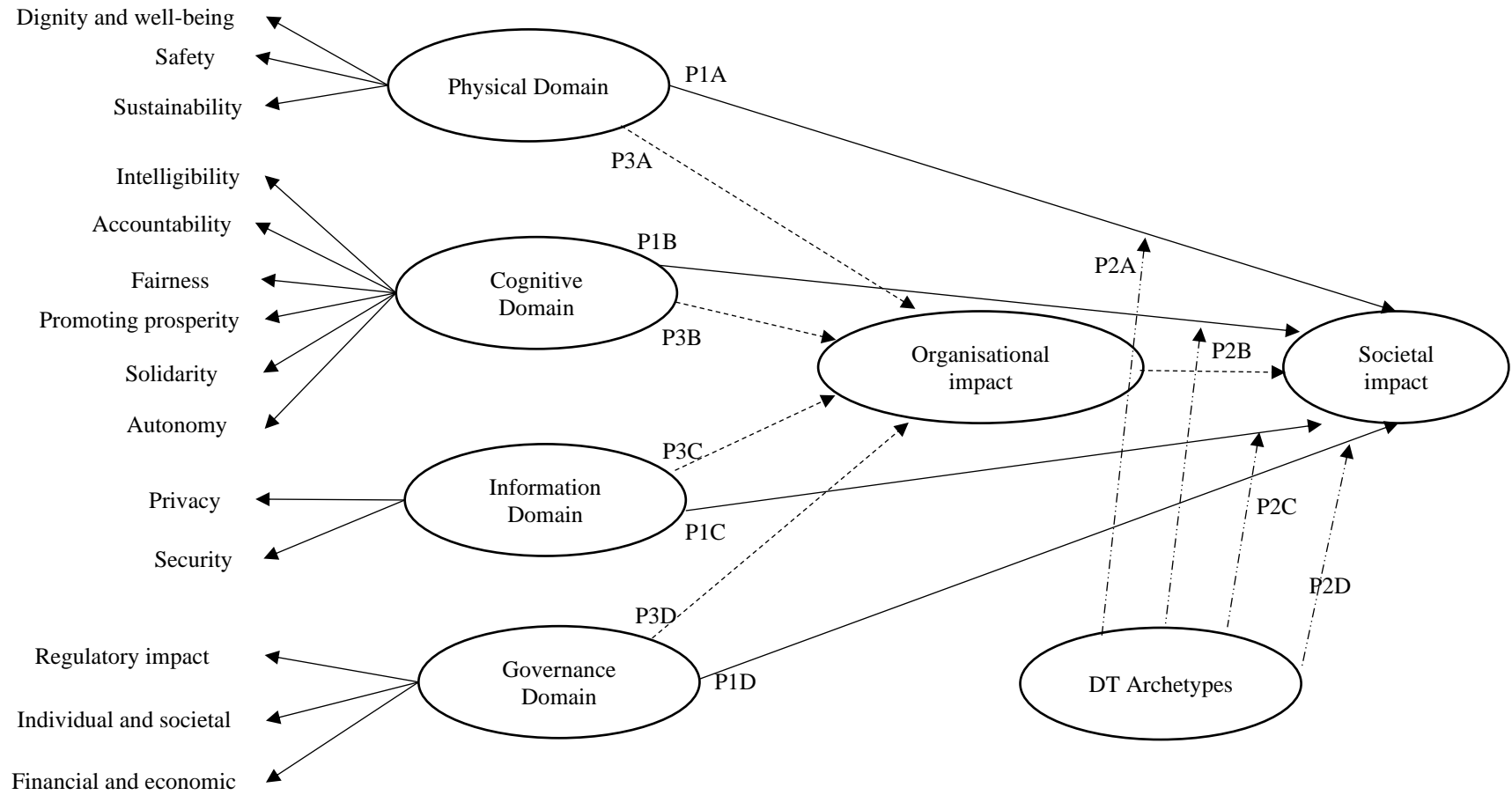
Figure 5: Conceptual Model for the digital ethical assessment of the use of AI in DTs

References

Adomavicius, G., Bockstedt, J. C., Gupta, A., & Kauffman, R. J. (2007). Technology roles and paths of influence in an ecosystem model of technology evolution. *Information Technology and Management*, *8*(2), 185-202.

Akter, S., Motamarri, S., Hani, U., Shams, R., Fernando, M., Babu, M. M., & Shen, K. N. (2020). Building dynamic service analytics capabilities for the digital marketplace. *Journal of Business Research*, *118*, 177-188. https://doi.org/10.1016/j.jbusres.2020.06.016

Anderson, M., & Anderson, S. L. (2006). Guest editors' introduction: Machine ethics. *IEEE Intelligent Systems*, *21*(4), 10-11.

Applin, S. (2017). Autonomous Vehicle Ethics: Stock or custom? *IEEE Consumer Electronics Magazine*, *6*(3), 108-110. https://doi.org/10.1109/MCE.2017.2684917

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.

Ashok, M. (2018). Role of Digitisation in Enabling Co-creation of Value in KIBS Firms. International Conference on Informatics and Semiotics in Organisations,

Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative Research*, *1*(3), 385-405. https://doi.org/10.1177/146879410100100307

Bain, R. (1937). Technology and State Government. *American sociological review*, *2*(6), 860-874. https://doi.org/10.2307/2084365

Barlow, C. (2020). Oncology Research: Clinical Trial Management Systems, Electronic Medical Record, and Artificial Intelligence. *Seminars in Oncology Nursing*, *36*(2), 151005. https://doi.org/10.1016/j.soncn.2020.151005

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press. https://books.google.ca/books?id=Q08oChJj8HQC

Berger, S., Denner, M.-S., & Roeglinger, M. (2018). The nature of digital technologies-development of a multi-layer taxonomy.

Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). *Hello, World! Artificial Intelligence and its Use in the Public Sector* Retrieved from https://www.oecd.org/governance/innovative-government/working-paper-hello-world-artificial-intelligence-and-its-use-in-the-public-sector.htm

Borges, A. F. S., Laurindo, F. J. B., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2020). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2020.102225

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey [https://doi.org/10.1002/asi.22803]. *Journal of the American Society for Information Science and Technology*, *64*(2), 217-233. https://doi.org/https://doi.org/10.1002/asi.22803

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence*, *1*, 316-334. https://doi.org/Book_Doi 10.1017/Cbo9781139046855

Breidbach, C. F., & Maglio, P. (2020). Accountable algorithms? The ethical implications of data-driven business models. *Journal of Service Management*, *31*(2), 163-185. https://doi.org/10.1108/josm-03-2019-0073

Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton.

Burk, D. L. (2019). Algorithmic Fair Use [Article]. *University of Chicago Law Review*, *86*(2), 283-307.

Buytendijk, F. (2019). Digital Ethics, or How to Not Mess Up With Technology, 2017. . *Gartner Report, ID: G00338664.*

Canada, G. o. (2020). *Algorithmic Impact Assessment*. Retrieved December 4, 2020 from https://open.canada.ca/aia-eia-js/?lang=en

Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, *63*(2), 183-193. https://doi.org/10.1016/j.bushor.2019.11.003

Cath, C., Zimmer, M., Lomborg, S., & Zevenbergen, B. (2018). Association of Internet Researchers (AoIR) Roundtable Summary: Artificial Intelligence and the Good Society Workshop Proceedings. *Philosophy and Technology*, *31*(1), 155-162. https://doi.org/10.1007/s13347-018-0304-8

Chen, P.-T., Lin, C.-L., & Wu, W.-N. (2020). Big data management in healthcare: Adoption challenges and implications. *International Journal of Information Management*, *53*. https://doi.org/10.1016/j.ijinfomgt.2020.102078

Chesbrough, H. (2010). Business Model Innovation: Opportunities and Barriers. *Long Range Planning*, *43*(2), 354-363. https://doi.org/https://doi.org/10.1016/j.lrp.2009.07.010

Cho, S., Vasarhelyi, M. A., Sun, T., & Zhang, C. (2020). Learning from Machine Learning in Accounting and Assurance. *Journal of Emerging Technologies in Accounting*, *17*(1), 1-10. https://doi.org/10.2308/jeta-10718

Conti, M., Gangwal, A., & Ruj, S. (2018). On the economic significance of ransomware campaigns: A Bitcoin transactions perspective. *Computers & Security*, *79*, 162-189. https://doi.org/10.1016/j.cose.2018.08.008

Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Denning, P. J., & Rosenbloom, P. S. (2009). The profession of IT Computing: the fourth great domain of science. *Communications of the ACM*, *52*(9), 27-29.

Dhagarra, D., Goswami, M., & Kumar, G. (2020). Impact of Trust and Privacy Concerns on Technology Acceptance in Healthcare: An Indian Perspective. *International Journal of Medical Informatics*, *141*. https://doi.org/10.1016/j.ijmedinf.2020.104164

Du, S., & Xie, C. (2020). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63-71. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., Le Meunier-FitzHugh, K., Le Meunier-FitzHugh, L. C., Misra, S., Mogaji, E., Sharma, S. K., Singh, J. B., Raghavan, V., Raman, R., Rana, N. P., Samothrakis, S., Spencer, J., Tamilmani, K., Tubadji, A., Walton, P., & Williams, M. D. (2021a). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for

research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021b). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, *59*. https://doi.org/10.1016/j.ijinfomgt.2020.102168

Ebert, I. (2019). The Tech Company Dilemma. Ethical Managerial Practice in Dealing with Government Data Requests [Article]. *Zeitschrift für Wirtschafts- und Unternehmensethik*, *20*(2), 264-275. https://doi.org/10.5771/1439-880x-2019-2-264

Farrell, J., & Weiser, P. J. (2003). Modularity, vertical integration, and open access policies: Towards a convergence of antitrust and regulation in the internt age. *Harv. JL & Tech.*, *17*, 85.

Fieser, J. (2000). *Metaethics, normative ethics, and applied ethics: historical and contemporary readings*. Wadsworth.

Flick, C., Zamani, E. D., Stahl, B. C., & Brem, A. (2020). The future of ICT for health and ageing: Unveiling ethical and social issues through horizon scanning foresight. *Technological Forecasting & Social Change*, *155*. https://doi.org/10.1016/j.techfore.2020.119995

Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, *376*(2133), Article 20180081. https://doi.org/10.1098/rsta.2018.0081

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, *1*(1).

Forum, W. E. (2020). *AI Procurement in a Box: Project overview*. Retrieved from https://www.weforum.org/reports/ai-procurement-in-a-box

Freeman, R. E. (1984). *Strategic management : a stakeholder approach*. Pitman.

Frizzo-Barker, J., Chow-White, P. A., Adams, P. R., Mentanko, J., Ha, D., & Green, S. (2020). Blockchain as a disruptive technology for business: A systematic review. *International Journal of Information Management*, *51*. https://doi.org/10.1016/j.ijinfomgt.2019.10.014

Gal, U., Jensen, T. B., & Stein, M.-K. (2020). Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization*, *30*(2). https://doi.org/10.1016/j.infoandorg.2020.100301

Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., Langer, S. G., Kitts, A. B., Birch, J., & Shields, W. F. (2019). Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Canadian Association of Radiologists Journal*, *70*(4), 329-334.

Gerlick, J. A., & Liozu, S. M. (2020). Ethical and legal considerations of artificial intelligence and algorithmic decision-making in personalized pricing. *Journal of Revenue and Pricing Management*, *19*(2), 85-98. https://doi.org/10.1057/s41272-019-00225-2

Giddens, A. (2018). A Magna Carta For The Digital Age [Article]. *New Perspectives Quarterly*, *35*(3), 6-8. https://doi.org/10.1111/npqu.12151

González-Rodríguez, M. R., Díaz-Fernández, M. C., & Pacheco Gómez, C. (2020). Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction

through emotions. *Telematics and Informatics*, *51*. https://doi.org/10.1016/j.tele.2020.101404

Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from# Brexit and# USElection. *European Economic Review*, 103772.

Gregory, A., & Halff, G. (2020). The damage done by big data-driven public relations. *Public Relations Review*, *46*(2). https://doi.org/10.1016/j.pubrev.2020.101902

Grewal, D., Kroschke, M., Mende, M., Roggeveen, A. L., & Scott, M. L. (2020). Frontline Cyborgs at Your Service: How Human Enhancement Technologies Affect Customer Experiences in Retail, Sales, and Service Settings. *Journal of Interactive Marketing*, *51*, 9-25. https://doi.org/10.1016/j.intmar.2020.03.001

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199-220. https://doi.org/https://doi.org/10.1006/knac.1993.1008

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, *51*(5), 1-42. https://doi.org/10.1145/3236009

Headrick, D. (2014). The Ethics and Law of Robots. *Research Technology Management*, *57*(3), 6-7.

Ioannou, A., Tussyadiah, I., & Lu, Y. (2020). Privacy concerns and disclosure of biometric and behavioral data for travel. *International Journal of Information Management*, *54*. https://doi.org/10.1016/j.ijinfomgt.2020.102122

ISO:26000. (2010). *Guidance on social responsibility*. ISO/FDIS 26000: 2010 (E), Final Draft, Geneva, November 1.

Janakiraman, R., Lim, J. H., & Rishika, R. (2018). The effect of a data breach announcement on customer behavior: Evidence from a multichannel retailer. *Journal of Marketing*, *82*(2), 85-105. https://doi.org/10.1509/jm.16.0124

Jaume-Palasi, L. (2019). Why We Are Failing to Understand the Societal Impact of Artificial Intelligence. *Social Research: An International Quarterly*, *86*(2), 477-498.

Joan, R. M. (2010). Information Ethics. *Online Dictionary For Library And Information Science. Np*.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

Johnson, D. G. (2015). Technology with No Human Responsibility? [Article]. *Journal of Business Ethics*, *127*(4), 707-715. https://doi.org/10.1007/s10551-014-2180-1

Jonas, H. (1985). Ethics and Biogenetic Art. *Social Research*, *52*(3), 491-504. http://www.jstor.org/stable/40970385

Kankanhalli, A., Charalabidis, Y., & Mellouli, S. (2019). IoT and AI for Smart Government: A Research Agenda. *Government Information Quarterly*, *36*(2), 304-309. https://doi.org/https://doi.org/10.1016/j.giq.2019.02.003

Karlsson, C. (2003). Spatial ICT clusters in Sweden-An empirical method to identify a necessary condition for existence.

Khalil, A., Ahmed, S. G., Khattak, A. M., & Al-Qirim, N. (2020). Investigating Bias in Facial Analysis Systems: A Systematic Review. *IEEE Access*, *8*, 130751-130761. https://doi.org/10.1109/Access.2020.3006051

Khan, S. A., Alkawaz, M. H., & Zangana, H. M. (2019). The use and abuse of social media for spreading fake news. 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS),

King, N. (2004). Using template analysis in the thematic analysis of text. In 'Essential guide to qualitative methods in organizational research'.(Eds G Symon, C Cassell) pp. 256–270. In. Sage: London.

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management science*, *65*(7), 2966-2981.

Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges [Article]. *Business Horizons*, *63*(2), 157-170. https://doi.org/10.1016/j.bushor.2019.10.005

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The Challenges of Algorithm-Based HR Decision-Making for Personal Integrity [Article]. *Journal of Business Ethics*, *160*(2), 377-392. https://doi.org/10.1007/s10551-019-04204-w

Liu, K. (2000). *Semiotics in information systems engineering*. Cambridge University Press.

Lodders, A., & Paterson, J. M. (2020). Scrutinising COVIDSafe: Frameworks for evaluating digital contact tracing technologies. *Alternative Law Journal*, *45*(3), 153-161, Article 1037969x20948262. https://doi.org/10.1177/1037969x20948262

Malkiat, T., Sarunkorn, C., & Stephen, J. (2019). Balancing information governance obligations when accessing social care data for collaborative research. *Records Management Journal*, *29*(1/2), 194-209. https://doi.org/10.1108/RMJ-09-2018-0029

Manfreda, A., Ljubi, K., & Groznik, A. (2021). Autonomous vehicles in the smart city era: An empirical study of adoption factors important for millennials. *International Journal of Information Management*, *58*. https://doi.org/10.1016/j.ijinfomgt.2019.102050

Martin, K. (2019a). Designing Ethical Algorithms. *MIS Quarterly Executive*, *18*(2), 129-142. https://doi.org/10.17705/2msqe.00012

Martin, K. (2019b). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*(4), 835-850. https://doi.org/10.1007/s10551-018-3921-3

Metcalf, J., Moss, E., & boyd, d. (2019). Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics. *Social Research*, *86*(2), 449-476.

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *Ai & Society*, *35*(4), 957-967. https://doi.org/10.1007/s00146-020-00950-y

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med*, *6*(7), e1000097.

Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, *16*(4), 266-275.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, *26*(4), 2141-2168. https://doi.org/10.1007/s11948-019-00165-5

Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*, *260*, 113172. https://doi.org/10.1016/j.socscimed.2020.113172

Mraović, B. (2008). Relevance of data mining for accounting: social implications. *Social Responsibility Journal*, *4*(4), 439-455. https://doi.org/10.1108/17471110810909858

Mulligan, D. K., & Bamberger, K. A. (2019). PROCUREMENT AS POLICY: ADMINISTRATIVE PROCESS FOR MACHINE LEARNING. *Berkeley Technology Law Journal*, *34*(3), 773-851. https://doi.org/10.15779/Z38RN30793

Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The Ethical Implications of Using Artificial Intelligence in Auditing. *Journal of Business Ethics*, *167*(2), 1-26. https://doi.org/10.1007/s10551-019-04407-1

Neubert, M. J., & Montañez, G. D. (2020). Virtue as a framework for the design and use of artificial intelligence. *Business Horizons*, *63*(2), 195-204. https://doi.org/10.1016/j.bushor.2019.11.001

Nicodemo, T. L., & Cardoso, O. (2019). Metahistory for (Ro)bots: Historical Knowledge in the Artificial Intelligence Era. *Historia da Historiografia*, *29*, 17-52. https://doi.org/10.15848/hh.v12i29.1443

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, *53*. https://doi.org/10.1016/j.ijinfomgt.2020.102104

O'Leary, D. E. (2019). GOOGLE'S Duplex: Pretending to be human [Article]. *Intelligent Systems in Accounting, Finance and Management*, *26*(1), 46-53. https://doi.org/10.1002/isaf.1443

O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., Holzinger, K., Holzinger, A., Sajid, M. I., & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *15*(1), e1968. https://doi.org/10.1002/rcs.1968

Ogden, C. R., & Richards, I. A. (1923). *The meaning of meaning*. IA.

Pathak, B., Ashok, M., & Tan, Y. L. (2020). Value co-destruction: Exploring the role of actors' opportunism in the B2B context. *International Journal of Information Management*, 102093.

Phillip, V. L. (1985). Defining 'Business Ethics': Like Nailing Jello to a Wall [research-article]. *Journal of Business Ethics*, *4*(5), 377.

Popper, K. (1979). *Three worlds*. University of Michigan.

Project, P. E., & Peirce, C. S. (1998). *The Essential Peirce, Volume 2*. Indiana University Press.

Raisch, S., & Krakowski, S. (2020). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *Academy of Management Review*.

Rhim, J., Lee, G.-b., & Lee, J.-H. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. *Computers in Human Behavior*, *102*, 39-56. https://doi.org/10.1016/j.chb.2019.08.010

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V. C., & Floridi, L. (2021). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *Ai & Society*, *36*(1), 59-77. https://doi.org/10.1007/s00146-020-00992-2

Robles Carrillo, M. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy*, *44*(6). https://doi.org/10.1016/j.telpol.2020.101937

Rosenbloom, P. S. (2012). *On computing: The fourth great scientific domain*. MIT Press.

Ruvinsky, A. I. (2007). Computational ethics. In *Encyclopedia of information ethics and security* (pp. 76-82). IGI Global.

Sarker, M. N. I., Wu, M., Cao, Q., Alam, G., & Li, D. (2019). Leveraging Digital Technology for Better Learning and Education: A Systematic Literature Review. *Int. J. Inf. Educ. Technol*, *9*(7), 453-461.

Savirimuthu, J. (2017). Do algorithms dream of 'data' without bodies? *International Review of Law, Computers & Technology*, *31*(2), 243-262. https://doi.org/10.1080/13600869.2017.1298498

Schappert, M., & von Hauff, M. (2020). Sustainable consumption in the smart grid: From key points to eco-routine. *Journal of Cleaner Production*, *267*. https://doi.org/10.1016/j.jclepro.2020.121585

Schniter, E., Shields, T. W., & Sznycer, D. (2020). Trust in humans and robots: Economically similar but emotionally different. *Journal of Economic Psychology*, *78*. https://doi.org/10.1016/j.joep.2020.102253

Shanmuganathan, M. (2020). Behavioural finance in an era of artificial intelligence: Longitudinal case study of robo-advisors in investment decisions. *Journal of Behavioral and Experimental Finance*, *27*. https://doi.org/10.1016/j.jbef.2020.100297

Siebecker, M. R. (2019). Making Corporations More Humane Through Artificial Intelligence. *Journal of Corporation Law*, *45*(1), 95-149.

Sipior, J. C. (2020). Considerations for development and use of AI in response to COVID-19. *International Journal of Information Management*, *55*, 102170. https://doi.org/10.1016/j.ijinfomgt.2020.102170

Smith, B., & Welty, C. (2001). Ontology: Towards a new synthesis. Formal Ontology in Information Systems,

Sousa, W. G. d., Melo, E. R. P. d., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, *36*(4), 101392. https://doi.org/10.1016/j.giq.2019.07.004

Stahl, B. C., Timmermans, J. O. B., & Mittelstadt, B. D. (2016). The Ethics of Computing: A Survey of the Computing-Oriented Literature. *ACM Computing Surveys*, *48*(4), 55:51-55:38. https://doi.org/10.1145/2871196

Subramanian, R. (2017). Emergent AI, Social Robots and the Law: Security, Privacy and Policy Issues. *Journal of International Technology & Information Management*, *26*(3), 81-105.

Sutton, S. G., Arnold, V., & Holt, M. (2018). How Much Automation Is Too Much? Keeping the Human Relevant in Knowledge Work. *Journal of Emerging Technologies in Accounting*, *15*(2), 15-25. https://doi.org/10.2308/jeta-52311

Tzafestas, S. G. (2018). Roboethics: Fundamental concepts and future prospects. *Information*, *9*(6), 148. https://doi.org/ARTN14810.3390/info9060148

Urquhart, L., Schnädelbach, H., & Jäger, N. (2019). Adaptive Architecture: regulating human building interaction. *International Review of Law, Computers & Technology*, *33*(1), 3-33. https://doi.org/10.1080/13600869.2019.1562605

Venturini, T., & Rogers, R. (2019). " API-based research" or how can digital sociology and journalism studies learn from the Cambridge Analytica data breach. *Digital Journalism*, *7*(4), 532-540.

Vesnic-Alujevic, L., Nascimento, S., & Pólvora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, *44*(6). https://doi.org/10.1016/j.telpol.2020.101961

Vidgen, R., Hindle, G., & Randolph, I. (2020). Exploring the ethical implications of business analytics with a business ethics canvas. *European Journal of Operational Research*, *281*(3), 491-501. https://doi.org/10.1016/j.ejor.2019.04.036

von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*.

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: a review and research agenda. *Journal of Database Management (JDM)*, *30*(1), 61-79. https://doi.org/10.4018/Jdm.2019010104

Watson, D. (2019). The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence. *Minds and Machines*, *29*(3), 417-440. https://doi.org/10.1007/s11023-019-09506-6

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. *London: Nuffield Foundation*.

Wiesenberg, M., & Tench, R. (2020). Deep strategic mediatization: Organizational leaders' knowledge and usage of social bots in an era of disinformation. *International Journal of Information Management*, *51*. https://doi.org/10.1016/j.ijinfomgt.2019.102042

Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons*, *61*(6), 823-832. https://doi.org/10.1016/j.bushor.2018.07.001

Xie, Y., Chen, K., & Guo, X. (2020). Online anthropomorphism and consumers' privacy concern: Moderating roles of need for interaction and social exclusion. *Journal of Retailing and Consumer Services*, *55*. https://doi.org/10.1016/j.jretconser.2020.102119

Yong, T., Jason, X., Rafael, B.-A., & Lakshmi, I. (2019). Ethics of blockchain : A framework of technology, applications, impacts, and research directions. *Information Technology & People*, *33*(2), 602-632. https://doi.org/10.1108/ITP-10-2018-0491

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research Commentary: The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, *21*(4), 724.

Appendix 1: Publications included in the literature review

	Publication
1.	Barlow, C. (2020). Oncology Research: Clinical Trial Management Systems, Electronic Medical Record, and Artificial Intelligence. *Seminars in Oncology Nursing, 36*(2).
2.	Breidbach, C. F., & Maglio, P. (2020). Accountable algorithms? The ethical implications of data-driven business models. *Journal of Service Management*.
3.	Burk, D. L. (2019). Algorithmic Fair Use. *University of Chicago Law Review, 86*(2), 283-307.
4.	Cath, C., Zimmer, M., Lomborg, S., & Zevenbergen, B. (2018). Association of Internet Researchers (AoIR) Roundtable Summary: Artificial Intelligence and the Good Society Workshop Proceedings. *Philosophy and Technology, 31*(1), 155-162.
5.	Chen, P.-T., Lin, C.-L., & Wu, W.-N. (2020). Big data management in healthcare: Adoption challenges and implications. *International Journal of Information Management, 53*.
6.	Cho, S., Vasarhelyi, M. A., Sun, T., & Zhang, C. (2020). Learning from Machine Learning in Accounting and Assurance. *Journal of Emerging Technologies in Accounting, 17*(1), 1-10.
7.	Dhagarra, D., Goswami, M., & Kumar, G. (2020). Impact of Trust and Privacy Concerns on Technology Acceptance in Healthcare: An Indian Perspective. *International Journal of Medical Informatics, 141*.
8.	Du, S., & Xie, C. (2020). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*.
9.	Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63-71.
10	Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021b). Setting the future of digital and social media marketing research: Perspectives and research propositions [Article]. *International Journal of Information Management, 59*.
11	Ebert, I. (2019). The Tech Company Dilemma: Ethical Managerial Practice in Dealing with Government Data Requests. *ZfWU - Journal for Business, Economics & Ethics, 20*(2), 264-275.
12	Flick, C., Zamani, E. D., Stahl, B. C., & Brem, A. (2020). The future of ICT for health and ageing: Unveiling ethical and social issues through horizon scanning foresight. *Technological Forecasting & Social Change, 155*.
13	Frizzo-Barker, J., Chow-White, P. A., Adams, P. R., Mentanko, J., Ha, D., & Green, S. (2020). Blockchain as a disruptive technology for business: A systematic review. *International Journal of Information Management, 51*.
14	Gal, U., Jensen, T. B., & Stein, M.-K. (2020). Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization, 30*(2).
15	Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., Langer, S. G., Kitts, A. B., Birch, J., & Shields, W. F. (2019). Ethics of artificial intelligence in

	Publication
	radiology: summary of the joint European and North American multisociety statement. *Canadian Association of Radiologists Journal*, 70(4), 329-334.
16	Gerlick, J. A., & Liozu, S. M. (2020). Ethical and legal considerations of artificial intelligence and algorithmic decision-making in personalized pricing. *Journal of Revenue and Pricing Management, 19*(2), 85-98.
17	Giddens, A. (2018). A Magna Carta For The Digital Age. *NPQ: New Perspectives Quarterly, 35*(3), 6-8.
18	González-Rodríguez, M. R., Díaz-Fernández, M. C., & Pacheco Gómez, C. (2020). Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. *Telematics and Informatics, 51*.
19	Gregory, A., & Halff, G. (2020). The damage done by big data-driven public relations. *Public Relations Review, 46*(2).
20	Grewal, D., Kroschke, M., Mende, M., Roggeveen, A. L., & Scott, M. L. (2020). Frontline Cyborgs at Your Service: How Human Enhancement Technologies Affect Customer Experiences in Retail, Sales, and Service Settings. *Journal of Interactive Marketing, 51*, 9-25.
21	Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys, 51*(5), 1-42.
22	Headrick, D. (2014). The Ethics and Law of Robots. *Research Technology Management, 57*(3), 6-7.
23	Ioannou, A., Tussyadiah, I., & Lu, Y. (2020). Privacy concerns and disclosure of biometric and behavioral data for travel. *International Journal of Information Management, 54*.
24	Johnson, D. (2015). Technology with No Human Responsibility? *Journal of Business Ethics, 127*(4), 707-715.
25	Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons, 63*(2), 157-170.
26	Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The Challenges of Algorithm-Based HR Decision-Making for Personal Integrity. *Journal of Business Ethics, 160*(2), 377-392.
27	Lodders, A., & Paterson, J. M. (2020). Scrutinising COVIDSafe: Frameworks for evaluating digital contact tracing technologies. *Alternative Law Journal*.
28	Malkiat, T., Sarunkorn, C., & Stephen, J. (2019). Balancing information governance obligations when accessing social care data for collaborative research. *Records Management Journal, 29*(1/2), 194-209.
29	Manfreda, A., Ljubi, K., & Groznik, A. (2019). Autonomous vehicles in the smart city era: An empirical study of adoption factors important for millennials. *International Journal of Information Management*.
30	Martin, K. (2019a). Designing Ethical Algorithms. *MIS Quarterly Executive, 18*(2), 129-142.
31	Martin, K. (2019b). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics, 160*(4), 835-850.
32	Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *Ai & Society*.

	Publication
33	Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine, 260*.
34	Mraović, B. (2008). Relevance of data mining for accounting: social implications. *Social Responsibility Journal*, *4*(4), 439-455.
35	Mulligan, D. K., & Bamberger, K. A. (2019). PROCUREMENT AS POLICY: ADMINISTRATIVE PROCESS FOR MACHINE LEARNING. *Berkeley Technology Law Journal, 34*(3), 773-851.
36	Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The Ethical Implications of Using Artificial Intelligence in Auditing. *Journal of Business Ethics*, 1-26.
37	Neubert, M. J., & Montañez, G. D. (2020). Virtue as a framework for the design and use of artificial intelligence. *Business Horizons, 63*(2), 195-204.
38	Nicodemo, T. L., & Cardoso, O. (2019). Metahistory for (Ro)bots: Historical Knowledge in the Artificial Intelligence Era. *Historia Da Historiografia, 29*, 17-52.
39	O'Leary, D. E. (2019). GOOGLE'S Duplex: Pretending to be human. *Intelligent Systems in Accounting, Finance & Management, 26*(1), 46-53.
40	O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., . . . Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery, 15*(1), e1968.
41	Rhim, J., Lee, G.-b., & Lee, J.-H. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. *Computers in Human Behavior, 102*, 39-56.
42	Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V. C., & Floridi, L. (2020). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *Ai & Society*.
43	Robles Carrillo, M. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy, 44*(6).
44	Savirimuthu, J. (2017). Do algorithms dream of 'data' without bodies? *International Review of Law, Computers & Technology, 31*(2), 243-262.
45	Schniter, E., Shields, T. W., & Sznycer, D. (2020). Trust in humans and robots: Economically similar but emotionally different. *Journal of Economic Psychology, 78*.
46	Shanmuganathan, M. (2020). Behavioural finance in an era of artificial intelligence: Longitudinal case study of robo-advisors in investment decisions. *Journal of Behavioral and Experimental Finance, 27*.
47	Siebecker, M. R. (2019). Making Corporations More Humane Through Artificial Intelligence. *Journal of Corporation Law, 45*(1), 95-149.
48	Sipior, J. C. (2020). Considerations for development and use of AI in response to COVID-19. *International Journal of Information Management, 55*.
49	Subramanian, R. (2017). Emergent AI, Social Robots and the Law: Security, Privacy and Policy Issues. *Journal of International Technology & Information Management, 26*(3), 81-105.

	Publication
50	Sutton, S. G., Arnold, V., & Holt, M. (2018). How Much Automation Is Too Much? Keeping the Human Relevant in Knowledge Work. *Journal of Emerging Technologies in Accounting, 15*(2), 15-25.
51	Urquhart, L., Schnädelbach, H., & Jäger, N. (2019). Adaptive Architecture: regulating human building interaction*. *International Review of Law, Computers & Technology, 33*(1), 3-33.
52	Vesnic-Alujevic, L., Nascimento, S., & Pólvora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy, 44*(6).
53	Vidgen, R., Hindle, G., & Randolph, I. (2020). Exploring the ethical implications of business analytics with a business ethics canvas. *European Journal of Operational Research, 281*(3), 491-501.
54	Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: a review and research agenda. *Journal of Database Management (JDM), 30*(1), 61-79.
55	Watson, D. (2019). The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence. *Minds and Machines, 29*(3), 417-440.
56	Wiesenberg, M., & Tench, R. (2020). Deep strategic mediatization: Organizational leaders' knowledge and usage of social bots in an era of disinformation. *International Journal of Information Management, 51*.
57	Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons, 61*(6), 823-832.
58	Xie, Y., Chen, K., & Guo, X. (2020). Online anthropomorphism and consumers' privacy concern: Moderating roles of need for interaction and social exclusion. *Journal of Retailing and Consumer Services, 55*.
59	Yong, T., Jason, X., Rafael, B.-A., & Lakshmi, I. (2019). Ethics of blockchain : A framework of technology, applications, impacts, and research directions. *Information Technology & People, 33*(2), 602-632.