

# South Korean Public Value Coproduction Towards ‘AI for Humanity’: A Synergy of Sociocultural Norms and Multistakeholder Deliberation in Bridging the Design and Implementation of National AI Ethics Guidelines

You Jeen Ha

hyujeen@gmail.com

Korea Advanced Institute of Science and Technology  
Daejeon, Republic of Korea

## ABSTRACT

As emerging technologies such as Big Data, Artificial Intelligence (AI), robotics, and the Internet of Things (IoT) pose fundamental challenges for global and domestic technological governance, the ‘Fourth Industrial Revolution’ (4IR) comes to the fore with AI as a frontrunner, generating discussions on the ethical elements of AI amongst key stakeholder groups, such as government, academia, industry, and civil society. However, in recent AI ethics and governance scholarship, AI ethics design appears to be divorced from AI ethics implementation, an implicit partition that results in two separate matters of theory and practice, respectively, and thus invokes efforts to bridge the ‘gap’ between the two. Such a partition potentially overcomplicates the discussion surrounding AI ethics and limits its productivity. This paper thus presents South Korea’s people-centered ‘National Guidelines for Artificial Intelligence Ethics’ (국가 인공지능 윤리기준; ‘Guidelines’) and their development under the Moon administration as a case study that can help readers conceptualize AI ethics design and implementation as a continuous process rather than a partitioned one.

From a public value perspective, the case study examines the Guidelines and the multistakeholder policymaking infrastructure that serves as the foundation for both the Guidelines’ design and implementation. This examination draws from literature in AI ethics and governance, public management and administration, and Korean policy and cultural studies as well as government and public documents alongside 9 interviews with members from the four stakeholder groups that collectively designed and continue to deliberate upon the Guidelines. Further, the study specifically focuses on (i) identifying public values that were highlighted by the Guidelines, (ii) investigating how such values reflect prevalent Korean sociocultural norms, and (iii) exploring how these values, in a way made possible by Korean sociocultural norms and policymaking, have been negotiated amongst the four stakeholder groups in a democratic public sphere to be ultimately incorporated into the

Guidelines and prepared for implementation. This paper hopes to contribute to theory-building in AI ethics and provide a point of comparison in the international stage for future research concerning AI ethics design and implementation.

## CCS CONCEPTS

• **Social and professional topics** → **Government technology policy**; *Governmental regulations*.

## KEYWORDS

4IR, Artificial intelligence, AI ethics, AI policy, Deliberative democracy, Multistakeholder, Networked governance, National Guidelines for Artificial Intelligence Ethics, Public value, Sociocultural norms, South Korea

## ACM Reference Format:

You Jeen Ha. 2022. South Korean Public Value Coproduction Towards ‘AI for Humanity’: A Synergy of Sociocultural Norms and Multistakeholder Deliberation in Bridging the Design and Implementation of National AI Ethics Guidelines. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’22)*, June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3531146.3533091>

## 1 INTRODUCTION

The ‘Fourth Industrial Revolution’ (hereafter referred to as ‘4IR’) or ‘Industry 4.0’<sup>1</sup> are becoming mainstream terms denoting a new assemblage of technological changes that pose fundamental challenges for global and domestic technological governance. Big Data, AI, robotics, and the Internet of Things (IoT) are some of the primary vehicles thought to enable such technological changes [36]. This paper understands artificial intelligence (AI) as the core technology driving 4IR [30, 34] and therefore a technology that thus far provides the best example showcasing policy challenges for emerging technologies in terms of the benefits and risks that they may potentially provide. As technologies such as AI become more pervasive in modern society, the intersection of 4IR and AI ethics becomes ever more relevant and pressing [44].

<sup>1</sup>Some use ‘4IR’ and ‘Industry 4.0’ interchangeably whereas others maintain that the two terms are distinct in their definition [75]. For this paper, I will lean toward the latter view, preferring the term ‘4IR’ as I agree with [75]’s distinction between the terms: that ‘4IR’ encompasses technological innovations as well as their “impact on civil society, governance structures, and human identity in addition to solely economic and manufacturing ramifications” whereas ‘Industry 4.0’ more specifically addresses technological advancements in manufacturing.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*FAccT ’22, June 21–24, 2022, Seoul, Republic of Korea*

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9352-2/22/06...\$15.00

<https://doi.org/10.1145/3531146.3533091>

A case study, as Lijphart [40] notes, can significantly contribute to the establishment of general propositions and thus theory-building in social science. On December 23, 2020, South Korea (hereafter Korea), being one of the most active countries in promoting the development and diffusion of AI, announced its ‘National Guidelines for AI Ethics’ (hereafter ‘the Guidelines’; 국가 인공지능 윤리기준) with the people-centered theme ‘AI for Humanity’ [26, 62]. Korea’s approach to the Guidelines’ development and their subsequent implementation under the Moon administration is a noteworthy case study, particularly given (1) notwithstanding the country’s understanding of AI as the driving force for 4IR rather than as one of many 4IR’s emerging technologies, its instrumentalist view of AI and consequent prioritization of humanity and human rights over AI [17, 34, 64]; (2) its status as a global ICT leader [23]; and (3) its current policymaking infrastructure where the national government drives AI ethics development by creating and mediating a public-good oriented platform for multistakeholder deliberation among those from public and private sectors. Moreover, taking into account Korea’s concurrent world-renowned response to the COVID-19 pandemic via its synergy of ICT infrastructure and multistakeholder partnerships [37, 68], these factors together enable Korea to navigate AI ethics policymaking with one foundational, ‘whole-of-society’ [38] infrastructure. By contrast, AI ethics initiatives and endeavors tend to be concentrated within the private sector for other countries such as the United States, leaving AI ethics in the hands of a small number of stakeholders and thus vulnerable to exclusionary politics perpetuated by the status quo [21]. Although the study presented in this paper focuses on only one country and, as such, has limited findings, examining Korea’s approach to the dynamic between AI ethics, diverse stakeholder groups, and public-good oriented policymaking mechanisms can nonetheless provide a point of comparison in the international stage as crises such as the pandemic expedite 4IR [13] and continue to raise questions for AI ethics implementation.

From a public value perspective, this paper examines Korea’s policymaking process for national AI ethics by (i) identifying public values that were highlighted by the Guidelines, (ii) investigating how such values reflect prevalent Korean sociocultural norms, and (iii) exploring how these values have been negotiated amongst the four stakeholder groups in a democratic public sphere to be incorporated into the Guidelines and prepared for implementation in a way made possible by Korean policymaking. Section 2 offers a theoretical review of academic literature in AI ethics (or AI governance), public administration and management, and Korean policy and cultural studies regarding public value, multistakeholder deliberation, and sociocultural norms with respect to policymaking and implementation of principles. This review introduces concepts that will be used to understand the synergistic relationship between public officials, publicly established values and beliefs, and public discourse surrounding national AI ethics in Korea, laying the foundation for Section 3, which presents a descriptive case study of South Korea’s state-led, participatory, multistakeholder initiative to create national AI ethics guidelines and a public good-oriented platform for continued discussion and debate. Section 4 discusses overall findings explored in the paper and their limitations. Section

5 summarizes the main arguments presented, emphasizing the potential impact the policymaking model in this paper might have in the international stage, and suggests future directions for research.

## 2 LITERATURE REVIEW

Literature in AI ethics and governance discusses a key problem that institutions in our information society are confronted with – namely how they can transition from designing AI ethics principles, guidelines, and codes to implementing these principles, guidelines, and codes – often framing the problem as a question of how to translate theory into practice with the assumption that the former is a matter substantially separate from the latter. Meanwhile, via a set of compatible interpretations of a concept called public value, literature in public management and administration provides an approach for analyzing institutions of governance and the policymaking processes they engage in to turn ideas into regulatory mechanisms. The role of public value in governance highlights the collaborative relationship between public officials, publicly established values and beliefs, and public discourse. Viewing the alleged problem of bridging theory and practice in AI ethics from a public value perspective is important as this paper takes public value to be inherent to both the design and implementation of AI ethics principles, guidelines, and codes; hence, public value reconfigures the two ostensibly discrete, separate realms of AI ethics into a continuous process. Finally, literature in Korean policy and cultural studies provides the context within which public value can be situated in terms of Korean sociocultural norms and policymaking, thereby setting up the discussion Korea’s policymaking process for national AI ethics from a public value perspective.

### 2.1 Bridging Theory and Practice in AI Ethics

Following the development of AI ethics guidelines and codes by multiple countries, private companies, and non-governmental organizations (NGOs), the task of practically implementing AI ethics has been rising to prominence as a pressing issue [21, 47, 79]. In recent years, scholars in AI ethics have been concerned with the question of how relevant institutions and stakeholders should translate the theory underlying AI ethics principles, guidelines, and codes into practice in the form of implementable technical and legal mechanisms [4, 14–16, 33, 47–49, 79]. This question assumes an inherent disjuncture between theory and practice. Whittlestone et al. [79] (p. 197), for instance, characterize the disjuncture as “the gap between abstract ethical principles and specific cases [that needs to be bridged]”, thus treating the theoretical basis of AI ethics principles, guidelines, and codes as a matter that is substantially separate from their implementation (hence the need to bridge the two). Further, [79] (p. 195) lament the limited nature of AI ethics principles and their utility as “a starting point”, calling for AI ethics as a field to instead focus more on “identifying and attempting to resolve the tensions that arise” when they are applied to specific sets of circumstances. While examining the tensions that emerge once AI ethics principles are applied would certainly be useful, could there be a perspective that can frame and conceptualize AI ethics design and implementation as a continuous process rather than a partitioned one?

## 2.2 Public Value

Upon facing the challenge of implementing AI ethics codes and guidelines, how may institutions such as government respond in terms of policy mechanisms and meet the needs of relevant stakeholder groups? One approach is the metric of policy legitimacy, which Hanberger [19] (p. 258) defines as “the product of satisfying felt needs and solving perceived problems.” Park et al. [67] (p. 322) emphasize that “in order to achieve [policy] legitimacy and policy success, it is thus vital [for the institution in question] to respond to the needs and problems recognized by diverse policy stakeholders”; more explicitly, stakeholders’ roles and subjective interpretations with respect to not only a given policy’s development process, but also the policy’s societal significance and impact are important to policy legitimacy [67].<sup>2</sup> Furthermore, Matti [42] views the degree of policy legitimacy as “the extent to which the substantive content of policy (i.e., its goals, strategies for reaching these goals, and motivations as to why these goals should be reached) can be justified by reference to publicly established values and beliefs”, using public opinion as an indicator of policy legitimacy. These views together point to the importance of stakeholder engagement from both public and private sectors in public discourse and how such engagement invokes public value as well as directly contributes to its creation [5, 25]. But what precisely is ‘public value’, especially with respect to policy legitimacy?

Mark H. Moore [46] coined the term ‘public value’ in order to delineate the role of public officials when it comes to democratic governance upon inquiring how public enterprises could improve their performance under the leadership of public executives and managers. [46] (pp. 293-294) critically discussed the need for a governance infrastructure whose design and maintenance are supported by public executives who “focus on increasing the value of the organizations they lead to the broader society [and thus ‘coproduce’ (i.e., collaboratively create) public value].” He expected these executives to do so by exercising their creativity and drive as well as being responsive to political feedback. Many interpretations of ‘public value’ emerged thereafter [1–3, 7, 8, 11, 46, 73, 74]. Among these were distinct yet potentially compatible interpretations, one such set of interpretations being the singular versus plural forms of the term [1]. While a number of scholars focused on public value (singular) as a broader idea rather than as individual principles, such as its role as an overarching governance paradigm (e.g., public value management (PVM) paradigm<sup>3</sup>) or a framework for measuring government performance, others such as Bozeman [8] directed their attention toward public values (plural), which they

<sup>2</sup>[66] (p. 638) undertake a similar project, interrogating the relationship between public interests, stakeholders, and policy legitimacy by taking what they term “a ‘stakeholder-oriented’ approach”:

“How can we detect whether a policy agenda proposed by political elites faithfully reflects the public’s demands or includes hidden intentions? To answer this question, this paper applies what we term a ‘stakeholder-oriented’ approach to the study of policy legitimacy, with a focus on the policy frames shaped by societal stakeholders.”

<sup>3</sup>[73] (p. 56) argues for the PVM paradigm as follows:

“[PVM] does offer a new paradigm and a different narrative of reform. Its strength lies in its redefinition of how to meet the challenges of efficiency, accountability, and equity and in its ability to point to a motivational force that does not rely on rules or incentives to drive public sector reform. It rests on a fuller and rounder vision of humanity than does either public administration or new public management. People are, it suggests, motivated by their involvement in networks and partnerships, that is, their relationships with others formed in the context of mutual respect and shared learning.”

understood as subjectively held norms or principles (e.g., liberty, equity, accountability, robustness, community, prosperity) [1, 6, 25]. If the interpretations of public value (singular and plural) are indeed compatible, the former as a governance paradigm enables a deliberative space for networks and partnerships,<sup>4</sup> encompassing the latter, which then are individual subjectively held principles that can be debated within said space.<sup>5</sup>

Most pertinent in terms of policy legitimacy and the arguably compatible interpretations of public value, nonetheless, appear to be the ones advanced by Benington [3] (p. 233, 235), Stoker [73] (p. 42, 47), and Bozeman [8] (p. 13). The three define ‘public value’ as such, respectively: (1) “public value [i]s part of a deliberative process, embedded within a democratic public sphere within which competing interests and contested values can be debated and negotiated [...] through a continuing process of dialogue”; (2) “the judgment of... public value is collectively built through deliberation involving elected and appointed government officials and key stakeholders [and] the governance of the public realm involves networks of deliberation and delivery in pursuit of public value”; and (3) “those providing normative consensus about (a) the rights, benefits, and prerogatives to which citizens should (and should not) be entitled; (b) the obligations of citizens to society, the state, and one another; and (c) the principles on which governments and policies should be based.” Benington’s definition recognizes the importance of creating public value via deliberation as a continuing process of debating and negotiating conflicting interests and values. Meanwhile, Stoker’s definition adds to that offered by Benington, also acknowledging the role of networks in presenting stakeholders’ conflicting interests and values to a democratic public sphere when governing the public realm. Bozeman’s interpretation diverges from the previous two by narrowing its scope of the term and taking a more normative stance; yet, Bozeman remains consistent with the two as his definition could easily refer to the normative ‘values’ that [3] and [73] (p. 43) both allude to. Bozeman’s pluralistic definition also make possible public value dilemmas, which Stone [74] (p. 14) points to, stating that it is the responsibility “of the political analyst [when evaluating policies]... to reveal and clarify the underlying value disputes so that people can see where they differ and move toward some reconciliation.”<sup>6</sup>

In summary, ‘public value’ as defined by Benington, Stoker, and Bozeman together *is created via networks of diverse stakeholders whose conflicting interests and normative public values are debated and negotiated through continued deliberation in pursuit of public value within a democratic public sphere.*<sup>7</sup> Torfing et al. [77] (pp.2-3) allude to this form of governance with their definition of interactive governance, which they understand as “the complex process through which a plurality of social and political actors with diverging interests interact in order to formulate, promote, and achieve

<sup>4</sup>See footnote 3.

<sup>5</sup>Acknowledging their compatibility, this paper will assign both of these interpretations equal weight and relevance in light of the case study presented.

<sup>6</sup>[74] (p. 14)’s discussion of value disputes and reconciliation parallels [79] (p. 195)’s view on “identifying and attempting to resolve the tensions that arise” when AI ethics principles are applied.

<sup>7</sup>Habermas et al. [18] (p. 49) elaborate on the concept of the ‘public sphere’, stating that it is “a realm of our social life in which something approaching public opinion can be formed. Access is granted to all citizens.” According to notes by Peter Hohendahl within [18]’s paper, their concept of the public sphere is made concrete through public participation.

common objectives by means of mobilizing, exchanging, and deploying a range of ideas, rules, and resources.”<sup>8</sup> Resting under the umbrella of interactive governance is a more specific kind of governance called networked governance. ‘Networked governance’, as described by [77] (p. 16), is “a horizontal articulation of interdependent but operationally autonomous actors who interact through negotiations that take place within a relatively institutionalised framework and facilitate self-regulated policymaking in the shadow of hierarchy.” Underscoring interdependence and negotiation, networked governance is perhaps closer in meaning to the ‘public value’ denoted by the confluence of interpretations advanced by Benington, Stoker, and Bozeman.

### 2.3 Top-Down Policymaking to Deliberative Democracy

If this paper is to examine Korea’s policymaking infrastructure from a public value perspective grounded in the literature previously discussed, then some sociopolitical context is necessary. Korean social norms are heavily influenced by tenets of Confucian ethics, including ‘the Three Bonds’ (3강): Ruler-Subject, Father-Son, and Husband-Wife [29, 55], and Korean social norms dictate the nature of human relationships, including but not limited to political hierarchies [78]. The fundamental structure of ‘The Three Bonds’ is rooted in “[t]he centrality of the father-son relationship”; this structure then “giv[es] added persuasive power to the political authority of the ruler and the husband” [78] (p. 123).<sup>9</sup> Accordingly, Korea has traditionally followed a Confucian sociopolitical model, one strongly characterized by state leadership, hierarchy, and collectivism [32, 71].

Following the consolidation of constitutional democracy in 1987, nevertheless, Korea’s top-down approach to policymaking shifted to a more multistakeholder model of policymaking while retaining its Confucian elements [66]:

“The policy process in South Korea... prior to the democratization of 1987 was mainly dominated by political elites, with societal stakeholders playing a much less crucial role. Democratization changed the political landscape from one of the elite dominations to one involving competition between state elites and societal stakeholders. Although societal stakeholders’

influence has increased with the process of democratization, political elites can still control more institutional resources because of the historical legacies of top-down governance (e.g., Lee and Kim 2019; Park et al. 2015; Ringen et al. 2011). The new relationship between the state and civil society in this more democratic context has thus been the key determinant in Korea’s policy process.”

Arguably, an intriguing political amalgam of tradition and modernity results, especially when considering the respective roles of the state and the governed society during the previous three administrations. The varied degrees of state authority relative to stakeholder involvement become apparent by observing simply two of the three administrations, as [66] note<sup>10</sup>:

“Given the strong presidential system of Korea, the president has always played a crucial role in the policy-making process (Ringen et al. 2011; Lee and Kim 2019). The administration of President Roh fostered democratic political processes such as citizen involvement and recruited more civil society activists and experts than previous administrations (e.g., Kim 2012; Lee and Yun 2011). By contrast, the Park administration mostly relied on an elitist approach to policy-making and adhered to conservative mainstream values (e.g., Moon 2016). Her administration was committed to a unilateral style of communication with societal actors, whereas the Roh administration was a much stronger proponent of two-way communication in the public sphere.”

According to four types of societal stakeholders, the degree of stakeholder involvement and engagement was perceived to be higher during the Roh administration than during the Park administration [66]:

“During the Park administration, four types of societal stakeholders (i.e., policy experts, lobbyists, conservative press, and progressive press) presented frames criticizing policy processes, including “lack of stakeholder participation in the policy process” and “criticism of government political attitudes.” These were not present in the Roh administration, which favored more inclusive and democratic policy-making. These policy frames reflected the complaints of societal stakeholders who were not allowed to participate in the policy processes under the conservative administration.”

Roh’s ideological successor and current president President Moon Jae-in and his administration have supported networked governance, increasing the level of stakeholder engagement in policymaking and policy implementation via a ‘whole-of-society’ approach, particularly for aid policy reform and especially the country’s response to the COVID-19 pandemic [37, 38, 50].

<sup>8</sup>[77] (p. 15) offer a more detailed breakdown of their definition here:

“This definition emphasizes three important features of interactive governance. First, interactive governance refers to a complex process rather than a more or less unified set of formal structures and institutions connected in a linear fashion. [...] Second, the process is driven by a collective ambition to define and pursue common objectives in the face of the presence of divergent interests and preferences. [...] Interdependency forces the public and private actors to interact in order to find joint solutions that solve emerging problems and exploit new opportunities and thus somehow contribute to the advancement of the interests of the different actors. Third, the process is decentered in the sense that common objectives are formulated and achieved in and through negotiated interaction among a plurality of actors from the State, the economy, and civil society. Hence, although governments often play a crucial role as facilitator and manager of policy interaction, there is no privileged center in public policymaking, but a number of competing actors and arenas, each of which contributes important resources, experiences, and ideas.”

<sup>9</sup>Since this paper will take a descriptive approach rather than an evaluative one to examining the relationship between Korean sociocultural norms, policymaking, and AI ethics development, an in-depth critique of whether cultural frameworks and the values they reproduce are appropriate for advancing public values such as equality will be left open for future research.

<sup>10</sup>This paper does not take an analysis of Korean politics and policymaking based solely on Confucianism as sufficient nor does it reduce Korea’s history to a mere linear transition from being ruled by an imperial, top-down regime to practicing governance as a modern democracy. This paper only discusses these particular aspects of Korean history in order to highlight certain sociocultural norms and a specific policymaking infrastructure in support of its case study.

Hence, this paper seeks to address the question of how a synergy of sociocultural norms such as those from Confucianism, multistakeholder deliberation, and public value coproduction for national AI ethics development in a democratic society such as Korea can contribute to the conceptualization of AI ethics implementation relative to their design.

### 3 CASE STUDY

#### 3.1 Data Collection

This paper develops a single descriptive case study to understand the development of national AI ethics in South Korea vis-à-vis the policymaking process that designed and produced the Guidelines. Data collection for this case study involved document-based research derivative from official government reports and publications and media articles (Appendix A) as well as 9 in-depth interviews with members from government, industry, academia, and citizen groups, the four key groups holding stakes in AI ethics development in Korea: a director of the Artificial Intelligence Policy Bureau in the Ministry of Science and ICT; a director of the government-affiliated research center Korea Information Society Development Institute (KISDI); two Working Group (WG) members from Korean legal and ethics scholarship who helped produce the initial and subsequent drafts of the 'National Guidelines for AI Ethics'; a director from Korea's Artificial Intelligence Industry Association; one computer scientist, one philosopher, and one legal scholar within Korean academia; and one digital human rights researcher who represents citizen groups (Appendix B). These interviews were used as a source of qualitative data to understand: (1) the stages of the Guidelines' development process, from the events leading up to the Guidelines' conception to the ethics checklists and public forums that followed the Guidelines' release; (2) the deliberative process in which the various stakeholder groups participated to negotiate public value dilemmas and ultimately coproduce public value for national AI ethics; and (3) the sociopolitical status quo and embedded cultural norms that laid the foundation for the Guidelines. The documents reviewed and the responses obtained from these interviews helped in the building of our case study from a public value perspective.

#### 3.2 Background

Following AlphaGo's victory against Go world champion Lee Sedol in 2016, heightened concerns regarding AI and its potential capabilities demanded further attention in the realm of ethics research and policy implementation [9, 39, 69, 70]. Of the numerous ethical issues resulting from AI systems around that year, one case involved Microsoft's AI chatbot Tay which shocked the American public when it quickly tweeted highly-offensive and racist language only a mere few hours after the bot began learning and emulating language patterns encountered through its interactions with Twitter users upon its release online [72]. Almost five years following the Tay incident, Korea encounters a similar issue with Scatter Lab's chatbot Lee Luda (이루다) in 2021. The chatbot, drawing from 10 billion real-life conversations among young couples from the country's most popular messaging app KakaoTalk, used language offensive to minorities, the disabled, pregnant women, and the LGBT community on Facebook, causing outrage among the Korean citizenry [35, 43]. While Scatter Lab initially adopted a defensive stance regarding

the chatbot, stating that the deep learning system merely needed more time for further training and education to stabilize, the developers ultimately suspended the service mid-January amid growing criticism from the public [35]. The AI systems' flaws as well as the developers' and the public's respective responses together highlight the importance of Korea's Guidelines and their role in establishing standardized norms for both the private and public sectors.

Korea's active involvement in AI ethics in terms of the Guidelines coincided more or less with the OECD Council's mobilization of the international AI Group of Experts at the OECD (AIGO) from 2018 to 2019. The efforts of AIGO, which constituted a High Level Expert Group's creation of a foundational framework and subsequent revisions based on numerous stakeholders' opinions [53], eventually led to the release of the Recommendation of the Council on Artificial Intelligence and thereafter the publication of the OECD Principles on Artificial Intelligence [52]. Korean government officials were active members of AIGO, with members of Korea's Ministry of Science and ICT as the AIGO Chair and Director of the AIGO's Multilateral Cooperation Division, respectively, and a member of Korea's National Information Society Agency as Executive Director of AIGO's Department of Big Data [53]. Our interviews revealed that the collaborative experience acquired from this initiative coupled with OECD's call for national governments to work with stakeholder groups to devise national AI policies consistent with its AI principles [53] acted as inspiration for a similar deliberative policymaking process for AI ethics at the intranational level in Korea. Thus, the development of Korea's own National AI Ethics Guidelines occurred via multistakeholder engagement and discourse.

Concurrent with Korea's considerable involvement with the OECD and AIGO was the country's increased investment in further fortifying and enhancing its ICT infrastructure in light of 4IR, especially after President Moon Jae-In took office in May 2017. Given Korea's world-renowned global ICT leadership, Moon underscored the importance of Korea's future role as a front-runner in 4IR; Moon and the Korean government thus initially set their sights for IoT to be one of the central pillars for Korea's endeavor toward the new technological revolution [41]. To further address these goals, President Moon launched the Presidential Committee on the Fourth Industrial Revolution (PCFIR; 대통령 직속 4차 산업혁명위원회) in the same year, tasking the presidential policy coordination body with (i) deliberating upon and coordinating policy measures set forth by government ministries and PCFIR members; (ii) organizing public campaigns on 4IR and encouraging public participation; (iii) laying the foundations for regulatory and institutional reforms in support of public-private partnerships; and (iv) fostering ecosystems for new industries through the Special Committee on Smart City, Special Committee on Healthcare, etc. [51]. With such policy goals and institutional initiatives in place, state-driven measures toward transitioning Korea from a global ICT leader into a global 4IR leader became more overt.

The following year, Korea garnered attention as a future-oriented 4IR leader at the 2018 Pyeongchang Olympics (also referred to as "Pyeongchang K-ICT Olympics") [27]. The country showcased its world-class, cutting-edge telecommunications advances, specifically its fiber optics network at work with 5G, virtual reality, robotics, AI-translation services, autonomous buses, and Internet of Things (IoT) with the "K-ICT" strategy

[12, 28, 31, 41, 54, 65]. Promotion of Korea's ICT services overseas and the consequent foreign media's reception of the Pyeongchang Olympics as "the most high-tech Olympics ever" required public-private partnerships among Korea's Ministry of Science and ICT and Ministry of Land, Infrastructure, and Transport and Korean tech companies such as Samsung Electronics, Korea Telecom (KT) Corporation, and Hyundai Motor Company as well as overseas companies such as Intel and Ericsson [28, 31, 65].

As public-private partnerships began to help foster 4IR, research on AI ethics at the local, corporate, and ultimately state level commenced more rigorously viz-à-viz, for instance, AI ethics charters (AICE) devised by parties such as Korean social media giant Kakao Corp, Korea Advanced Institute of Science and Technology (KAIST), the National Information Society Agency, the Korea Artificial Intelligence Ethics Association, the Korea Information Society Development Institute (KISDI), and the Ministry of Science and ICT [17, 22]. With the exception of the Ministry of Commerce Industry and Energy's 2007 Draft of the Robot Ethics Charter (DREC), the AICEs devised by the aforementioned parties were released in 2018-2019 [22]. [17] (p.607) notes that the attitude dictating Korea's approach to AI ethics is instrumentalist (i.e., tool-based), pointing to "its establishment of a clear human-over-machine hierarchy, where humans are of the highest priority and AI and robots are expected to support and further enhance this position of dominance." Korea's instrumentalist attitude toward AI takes center stage at the national level as the Korean government narrows its focus from 4IR to AI from 2019 onward.

President Moon and the Korean government reemphasize their aims to establish Korea as a front-runner in 4IR from 2017, except this time with AI. Moon announced the 'Presidential Initiative for AI' in October 2019 to increase public awareness and accumulate national resources regarding AI for the government has designated AI as "the decisive driver of the 4th Industrial Revolution"; in December 2019, Moon, together with the Korean government, released AI policy framework *Toward AI World Leader beyond AI: National Strategy for Artificial Intelligence* in December 2019 [64]. The AI ethics portion of the policy framework states the importance of "realiz[ing] people-centered AI [by] establish[ing] a global level of AI code of ethics based on social debate and consensus" [64].<sup>11</sup> In 2020, the focus on people-centered AI becomes clear with the Guidelines' 'AI for Humanity' theme and the Confucian values embedded within, and the emphasis on social debate and consensus to realize this philosophy manifests throughout and after the Guidelines' development process.

### 3.3 State-Driven AI Ethics Development: the 'National Guidelines for Artificial Intelligence Ethics'

With the aforementioned goals, initiatives, and public-private partnerships established, the Korean government began developing national AI ethics for the country's citizens and tech companies through the design and production of the Guidelines [62]. While East Asian governments such as that of Korea are typically known

for their primarily top-down policy paradigms [45], the development of these Guidelines did not adhere to such a paradigm. The Guidelines were produced through a hybrid top-down and bottom-up approach, bringing together government, academia, industry, and civil society in a multistakeholder engagement grounded in public discourse, negotiation, and consensus. The Korean government only tasked the Ministry of Science & ICT and government-affiliated research institute KISDI with mediating the Guidelines' development process in 2019.<sup>12</sup> Revisiting the 'whole-of-society' approach utilized for other ventures [38, 50], the state created a space for the coproduction of public value (singular) for national AI ethics. Interviewees B and D indicated that the Korean government in joint collaboration with both the Ministry of Science and ICT and KISDI thereafter began assembling multiple AI ethics guidelines and principles documents produced intra- and internationally (such as the Korean AICEs), launching the Guidelines' public value coproduction process.

Beginning in April 2020, KISDI called for a Working Group (WG) of Korean legal, ethics, and ethics education scholars to devise the foundational theoretical framework for the Guidelines as these academics not only had the most experience researching technology ethics in the country, but also were deemed most directly pertinent to ethics development. The WG, according to interviewee I, created this framework in "zero-base" circumstances as there was no particular national-level precedent to draw upon; thus, from August to September 2020, the initial draft of the Guidelines was devised in a top-down manner. Subsequently, the views of 70 experts from academia (scholars in computer science, electrical engineering, law, ethics, medical law and ethics, humanities, and other AI-related fields), industry (developers, service distributors), and civil society (consumer groups) regarding this version were gathered via online discussion platforms and surveys for the collective bottom-up revision of the draft. Interviewee B noted that all stakeholders involved agreed on the overarching theme 'AI for Humanity' with 'public good' as the main normative public value (as defined by [8] (p. 13)).<sup>13</sup> As different viewpoints inevitably surfaced, however, a number of other key public values (plural) and eventually public value dilemmas were negotiated and deliberated upon with the WG's elementary framework as the point of departure; the 70 experts offered specialized corrections and suggestions to make it more holistic, inclusive, accessible, and multidisciplinary.

A fundamental change brought about by these stakeholder groups through this deliberative process, as noted by the majority of the interviewees, was the shift from a more traditional Confucian-laden value framework to a linguistically more modern and easily accessible (by multiple generations and cultures) value framework, which took the form of '3 Main Principles and 10 Key Requirements'

<sup>11</sup> 'People-centered' and 'human-centered' appear to be used interchangeably in government documents and press releases, so this paper will do the same.

<sup>12</sup>With respect to the Korean government's role in the Guidelines' development process, [77] (p. 15) definition of interactive governance is especially relevant: "Hence, although governments often play a crucial role as facilitator and manager of policy interaction, there is no privileged center in public policymaking, but a number of competing actors and arenas, each of which contributes important resources, experiences, and ideas."

<sup>13</sup>'Public good' seemed to be viewed as the equivalent of 'For Humanity' by stakeholder groups.

(3대 원칙과 10대 요건).<sup>14</sup> Each of the 3 Main Principles (i.e., Respect for Human Dignity (인간 존엄성), Consensus Good of Society (사회의 공공선), Fitness for Purpose (기술의 합목적성)) places human beings and the welfare of humanity "from a social, national, and global perspective" above AI technologies, which the 3rd Main Principle specifically defines as merely "tools for humanity and human life" [62]. The 3 Main Principles, explains interviewee I, are nonetheless "derivative of the Confucian values previously incorporated into the framework," most specifically the three-pronged structure of the 'Three Bonds' and their hierarchical nature which places one party superior to another. Upon inspecting the 3 Main Principles, it becomes evident that 'AI for Humanity' is an appropriate theme for the Guidelines as each principle places humans superior to AI systems; this humanity-prioritizing dynamic embodies the instrumentalist view of AI that is characteristic of previous local and corporate Korean AI ethics guidelines [17]. From a general point of view, the main public value under negotiation in this instance was accessibility with a global demographic in mind, and the public value dilemma concerned accessibility versus a lack thereof in favor of preserving cultural norms in their traditional form. A compromise was then reached by simplifying the potentially esoteric 'Three Bonds' to their structure and hierarchical essence without sacrificing comprehensibility. Furthermore, the interfusion of Confucian ethics and instrumentalist attitudes in the Guidelines via this deliberative process exemplifies the kind of public value creation that [3, 8, 73]'s definitions together refer to (see Section 2.2).

Enveloped by the humanity-prioritizing, instrumentalist 3 Main Principles are the '10 Key Requirements', namely Human Rights (인권보장), Protection of Privacy (프라이버시 보호), Respect for Diversity (다양성 존중), Prevention of Harm (침해금지), Public Good (공공성), Solidarity (연대성), Data Management (데이터 관리), Accountability (책임성), Safety (안전성), and Transparency (투명성), in no particular order. The Requirements' interpretation and implementation are expected to align with the 3 Main Principles. Based on the Ministry of Science and ICT of Korea [57]'s report and the interviews for this study, another substantial public value dilemma under negotiation among academia, industry, and civil society involved two of the 10 Key Requirements: security (i.e., Protection of Privacy) versus transparency. While citizens' groups demanded greater transparency at the expense of secure AI systems, emphasizing the need for public understanding of these systems, academic and industry pushed back due to intellectual property and data security concerns as well as AI systems' technical limitations as 'black boxes.' As the degree of transparency desired by civil society was simply unfeasible due to currently unresolvable problems, the Ministry of Science and ICT and KISDI deferred to the judgment of those in academia and industry. Interviewee C expressed her discontent with this decision on behalf of the citizens' groups, inquiring whether it gave equal weight to the views of each stakeholder group involved. On the other hand, interviewee D's

account revealed that citizens' groups did not provide a response when asked for further thoughts during the deliberation process.

Due to her technical background, interviewee A from academia framed this dilemma a bit differently, speaking more to privacy and accuracy relative to transparency. Framing the value dispute instead as a three-way trade-off among privacy, accuracy, and transparency, she commented: "Any time you try to make AI more accurate, you require more specific, detailed data [i.e., infringe on privacy]. That means either the model becomes a black box that nobody can understand [i.e., lack of transparency], or if you do try to understand the output, you have to look at all the data points [i.e., increase transparency], which results in the infringement of privacy rights." Either trade-off and how each was deliberated upon among the stakeholder groups elucidate not only their conflicting interests and values, an aspect of public value highlighted by [3], but also the difficult, complex nature of interactive and networked governance due to their unavoidable entanglement with subjectivity.

More views from citizens were collected for approximately another month before the Guidelines' official release, an effort that involved a public hearing on December 7, 2020 where the Ministry of Science and ICT introduced the draft of the Guidelines [58]. The approximately 2-hour-long public hearing was opened by KISDI Director Moon Jung Wook who presented a summary of the Guidelines' development process and content. Subsequently, 10 other experts from the four stakeholder groups participated in another deliberative process, albeit this time in a publicly available (both in-person and on YouTube) discussion panel [63]. The 10 experts provided their respective views regarding the Guidelines, which ranged from their praise for the human-centered 'AI For Humanity' theme to their concerns about the abstract, general nature of the Principles and Key Requirements and how institutions could translate the ethics guidelines into policy and law.<sup>15</sup> Serving as a visual demonstration of public discourse, the public hearing ended while once more stressing the need for continued dialogue after the Guidelines' release in terms of their interpretation and implementation by means of policy mechanisms and legislation.

The Korean government officially released the finalized version of the Guidelines for public viewing two weeks later on December 23, 2020 [57]. As their ultimate form depended on the examination and approval of the Government à la the PCFIR [58], their development process concluded with a top-down approach, as affirmed by interviewee H. However, as highlighted both in the document itself and in the December public hearing, central to the Guidelines' societal role is their invocation of continued public discourse and discussion among stakeholders; the Guidelines consequently act as a springboard for Korea's implementation of AI ethics. [62] makes this role clear:

"Plans and strategies to implement this guideline will be prepared. This includes using the guideline as a platform to promote discussions on issues related to AI ethics among diverse stakeholders and developing specific checklists for involved parties through continued discussions and deliberations."

<sup>14</sup>Formed on the basis of 'soft law', as described by interviewee I, and voluntary norms, Korea's 'National Guidelines for AI Ethics' is encompassed by the overarching theme 'AI For Humanity' (인간성을 위한 인공지능) and structured as '3 Main Principles and 10 Key Requirements' (3대 원칙과 10대 요건). The Guidelines are thus non-binding, allowing for different interpretations - an important element of creating public value [3].

<sup>15</sup>Interviewees A and F expressed their concerns regarding the feasibility of this translation effort, echoing the views of the AI ethics scholars I include in the literature review.

Indeed, after the Guidelines' release, legislative hearings and policy seminars were held for further public discourse and discussion among experts from the four stakeholder groups, who then began inquired about interpretability and the feasibility of enforcing the Guidelines via technical and legal mechanisms [24, 60, 61]. Interviewee G confirmed that KISDI has been developing ethics checklists based on the Guidelines for both industry and consumers while those in industry have been focusing on how to verify the trustworthiness of AI systems. New government documents indeed have been produced, taking the Guidelines as the starting point for the development of more detailed and specific applications. Such documents include the 'Plan for Implementation of Human-Centered, Trustworthy AI', the 'AI Personal Information Protection Self-Checklist (for Developer and User Use)', and the 'Basic Principles for Protection of Users of AI-Based Media Recommendation Services' [10, 56, 59]. The 'AI Personal Information Protection Self-Checklist (for Developer and User Use)' in particular extracted 6 main principles from the Guidelines, 개인정보보호 중심설계 (Privacy by Design), and Korea's Personal Information Protection Act (PIPA), thereby creating a fusion of technical, legal, and ethical principles – yet an example of a multistakeholder, multidisciplinary effort to implement the theory underlying the Guidelines [56].

Thus, in Korea's case, [79]'s (p. 195) call for AI ethics to increase its practical utility as a field by focusing more on "identifying and attempting to resolve the tensions that arise" when AI ethics principles are applied to specific instances is already anticipated and addressed in the Guidelines. In this respect, public value has had a consistent presence both during and following the Guidelines' development process: public value (singular and plural) coproduction by four key stakeholder groups not only established the theoretical underpinnings of the Guidelines, but also continues to be a fundamental staple of policymaking for AI ethics as Korean institutions investigate how to implement the Guidelines. Korea's policymaking infrastructure has enabled this consistent presence and, as such, can frame and conceptualize AI ethics design and implementation as a continuous process without necessarily uncoupling the two into separate problems to address.

## 4 DISCUSSION

This paper seeks to present a descriptive case study of South Korea's state-led initiative to create national AI ethics guidelines which can help readers and thinkers conceptualize AI ethics design and implementation as a continuous process without necessarily partitioning the two into separate problems with a gap in between. Literature review as well as qualitative analysis of relevant government and public documents and interviews show that South Korea's approach to national AI ethics is a synergy of humanity-prioritizing, instrumentalist norms, Confucian ethics, multistakeholder deliberation, and public value coproduction facilitated by a recent history of networked governance within a Habermasian democratic public sphere. From a public value perspective, the case study demonstrates how particular sociocultural norms and values inform the definition and implementation of AI ethics principles. Further, the country's state-driven, yet multistakeholder policymaking approach to developing national AI ethics shows how AI ethics design and implementation can be treated as a part of one continuous process via public policy, similar to how public value can be defined as a part of continuous

dialogue and deliberation. Nonetheless, the qualitative data from this case study could be bolstered or reinterpreted with a larger sample of testimonies (which could not be obtained initially due to time constraints), given that 70 experts' views were consulted during the Guidelines' development. This sample may reveal new observations regarding the public values prioritized or public value dilemmas negotiated among the four stakeholder groups. Although Korea's case is one example, Korea's context-specific, participatory, multidisciplinary, and multistakeholder deliberative process could be illuminating for other countries, such as those that delegate AI ethics to a small group of actors with little coordination among relevant stakeholder groups. The case study presented in this paper thus raises additional questions that merit further research:

- In 5-10 years, from an evaluative perspective, could the multistakeholder deliberative process examined in this case study be deemed a success, and according to what particular metrics?
- How might ordinary citizens in Korea perceive the national government's initiative regarding AI ethics?
- What might comparisons between this case study in Korea and other countries' policymaking approaches to AI ethics reveal?
- As seen with the case study in this paper, how can ethical frameworks for AI respond to specific cultural values and sociocultural contexts? How would these frameworks relate to internationally established ethical frameworks for AI, and could frameworks developed at different levels (e.g., national versus international) successfully coexist and complement one another?

## 5 CONCLUSION

This paper examines Korea's policymaking process for designing and implementing people-centered 'National Guidelines for Artificial Intelligence Ethics' (국가 인공지능 윤리기준) from a public value perspective, seeking to understand how Korea approaches the problem of translating AI ethics design into AI ethics implementation as a part of public policy. Specifically, this paper investigates the relationship between the public values prioritized by the Guidelines, how such values reflect prevalent Korean sociocultural norms found in Confucian ethics, and how these values were negotiated amongst the four stakeholder groups involved in the deliberation process (i.e., government, academia, industry, and civil society). This investigation finds that public value, multistakeholder deliberation, and Korean sociocultural norms with respect to policymaking and implementation of principles together can explain how public officials, publicly established values and beliefs, and participatory discourse enabled state-led national AI ethics development in a country that already deems AI as the driving force for 4IR, prioritizes humanity and human rights over AI, enjoys its status as a global ICT leader, and responds to global crises via a collaborative fusion of high-end ICT infrastructure and multistakeholder partnerships. Moving forward, Korea's policymaking model for national AI ethics has the potential to serve as a useful model for other countries that not only are tackling the problem of integrating AI ethics design and implementation, but also wish to pursue a more publicly engaging means of developing AI ethics.



## ACKNOWLEDGMENTS

**Funding/Support:** This research was supported by the Korean-American Educational Commission (KAEC; also Fulbright Korea) through a Fulbright U.S. Student Program Research Grant Award with Korea Advanced Institute of Science and Technology (KAIST) as the author's host institution. The KAEC had no role in the design and conduct of the study; access and collection of data; analysis and interpretation of data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. Korea Advanced Institute of Science and Technology (KAIST)'s Korea Policy Center for the Fourth Industrial Revolution (KPC4IR) facilitated the design and conduct of the study as well as access to data and its collection but had no role in analysis and interpretation of data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. The author declares no other financial interests.

**Institutional Review Board Statement:** The investigation adhered to the ethics of Social-Behavioral-Educational Human Subjects Research and was approved by the KAIST Institutional Review Board (KH2021-076). Written consent was obtained from the interviewees after they were given an explanation about the study's purpose and data collection procedure.

**Informed Consent Statement:** Written consent was obtained from the interviewees after they were given an explanation about the study's purpose and data collection procedure.

**Data Privacy Statement:** Interviewee identities have been kept confidential to protect their privacy.

**Acknowledgments:** I would like to thank the Korean-American Educational Commission (KAEC) for its support of my independent research project during my grant period in 2021. I would also like to thank Professor So Young Kim (KPC4IR) and Dr. Moonjung Yim (formerly KPC4IR) for their help with access to resources for data collection (including the campus library). I am grateful to the 9 interviewees who offered their time and shared their diverse views for the qualitative portion of this paper. Lastly, I acknowledge the work of both the experts in South Korean government, academia, industry, and civil society who are committed to researching the importance of AI ethics and its role in protecting the well-being of not only Korean citizens, but the global community at large and those working to share knowledge in this interdisciplinary junction.

## REFERENCES

- [1] John Alford and Janine O'Flynn. 2009. Making Sense of Public Value: Concepts, Critiques and Emergent Meanings. *International Journal of Public Administration* 32, 3-4 (March 2009), 171–191. <https://doi.org/10.1080/01900690902732731>
- [2] Frank Bannister and Regina Connolly. 2014. ICT, public values and transformative government: A framework and programme for research. *Government Information Quarterly* 31, 1 (Jan. 2014), 119–128. <https://doi.org/10.1016/j.giq.2013.06.002>
- [3] John Benington. 2009. Creating the Public In Order To Create Public Value? *International Journal of Public Administration* 32, 3-4 (March 2009), 232–249. <https://doi.org/10.1080/01900690902749578>
- [4] Reuben Binns. 2018. Fairness in Machine Learning: Lessons from Political Philosophy. In *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency (Conference on Fairness, Accountability, and Transparency (ACM FAccT))*. 149–159. <https://proceedings.mlr.press/v81/binns18a.html>
- [5] James Bohman. 2000. *Public Deliberation: Pluralism, Complexity, and Democracy* (reprint ed.). MIT Press, Cambridge, MA.
- [6] Phillip Boyle. 2001. Public Problems, Values, and Choices. *Popular Government* (2001). <https://api.semanticscholar.org/CorpusID:140784344>
- [7] Barry Bozeman. 2002. Public-Value Failure: When Efficient Markets May Not Do? *Public Administration Review* 62, 2 (Jan. 2002), 145–161. <https://doi.org/10.1111/0033-3352.00165>
- [8] Barry Bozeman. 2007. *Public Values and Public Interest: Counterbalancing Economic Individualism*. Georgetown University Press, Georgetown, Washington DC.
- [9] Roger Brownword. 2017. From Erehwon to AlphaGo: for the sake of human dignity, should we destroy the machines? *Law, Innovation and Technology* 9, 1 (Jan. 2017), 117–153. <https://doi.org/10.1080/17579961.2017.1303927>
- [10] Korea Communications Commission. 2021. Basic Principles for Protection of Users of AI-Based Media Recommendation Services. Document. Retrieved September 28, 2021 from <https://www.kcc.go.kr/download.do?fileSeq=51935>
- [11] Antonio Cordella and Leslie Wilcocks. 2010. Outsourcing, bureaucracy and public value: Reappraising the notion of the 'contract state'. *Government Information Quarterly* 27, 1 (Jan. 2010), 82–88. <https://doi.org/10.1016/j.giq.2009.08.004>
- [12] Susan Crawford. 2018. *Fiber: The Coming Tech Revolution – and Why America Might Miss It*. Yale University Press, New Haven, CT.
- [13] Ruy de Castro Sobrosa Neto, Janayna Sobrosa Maia, Samara de Silva Neiva, Michael Dillon Scalia, and José Baltazar Salgueirinho Osório de Andrade Guerra. 2020. The fourth industrial revolution and the coronavirus: a new era catalyzed by a virus. *Research in Globalization* 2 (Dec. 2020). <https://doi.org/10.1016/j.resglo.2020.100024>
- [14] Sina Fazelpour and Zachary C. Lipton. 2020. Algorithmic Fairness from a Non-Ideal Perspective. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AI/ES '20)*. Association for Computing Machinery, New York, NY, USA, 149–159. <https://doi.org/10.1145/3375627.3375828>
- [15] Luciano Floridi. 2017. Infraethics—on the Conditions of Possibility of Morality. *Philosophy and Technology* 30, 4 (Dec. 2017). <https://doi.org/10.1007/s13347-017-0291-1>
- [16] Luciano Floridi. 2019. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology* 32 (2019), 185–193. <https://doi.org/10.1007/s13347-019-00354-x>
- [17] Danit Gal. 2020. Perspectives and Approaches in AI Ethics: East Asia. In *The Oxford Handbook of Ethics of AI*, Markus D. Dubber, Frank Pasquale, and Sunit Das (Eds.). Oxford University Press, Oxford, 606–624. <https://doi.org/10.1093/oxfordhb/9780190067397.013.39>
- [18] Jürgen Habermas, Sara Lennox, and Frank Lennox. 1974. The Public Sphere: An Encyclopedia Article. *New German Critique* 3 (1974), 49–55. <https://doi.org/10.2307/487737>
- [19] Anders Hanberger. 2003. Public policy and legitimacy: A historical policy analysis of the interplay of public policy and legitimacy. *Policy Sciences* 36, 3 (Dec. 2003), 257–278. <https://doi.org/10.1023/B:OLIC.0000017471.88620.9a>
- [20] Vincent M. Hendricks, Peter Blanken, and Nico F.P. Adriaans. 1992. *Theoretical and practical considerations*. IVO, Rotterdam, Netherlands.
- [21] Merve Hickok. 2021. Lessons learned from AI ethics principles for future actions. *AI and Ethics* 1, 1 (Feb. 2021), 41–47. <https://doi.org/10.1007/s43681-020-00008-1>
- [22] Ha Hwang and Min-Hye Park. 2020. The Threat of AI and Our Response: The AI Charter of Ethics in South Korea. *Asian Journal of Innovation and Policy* 9, 1 (April 2020), 56–78. <https://doi.org/10.7545/AJIP.2020.9.1.056>
- [23] U.S. Department of Commerce International Trade Administration. 2021. *South Korea - Information and Communication Technology*. Retrieved August 29, 2021 from <http://www.trade.gov/country-commercial-guides/south-korea-information-and-communication-technology>
- [24] Pilmo Jung. 2021. Legislative Hearing for Act on Creating Foundation for Trust and Fostering Artificial Intelligence. Video. Retrieved September 23, 2021 from <https://www.youtube.com/watch?v=n-Ch4-VWrc8>
- [25] Torben Beck Jørgensen and Barry Bozeman. 2007. Public Values: An Inventory. *Administration & Society* 39, 3 (May 2007), 354–381. <https://doi.org/10.1177/0095399707300703>
- [26] Hyeong-Woo Kan. 2020. *Presidential committee lays out first ethical standards for AI*. Retrieved September 21, 2021 from <http://www.koreaherald.com/view.php?ud=20201223000794>
- [27] Hyongsun Kim and Eunyoung Lee. 2015. A Study on ICT Strategies of PyeongChang Winter Olympic Games. In *Proceedings of the Korea Information Processing Society Conference (Korea Information Processing Society Conference, Vol. 3)*. Korea Information Processing Society, 245–246. <https://doi.org/10.3745/PKIPS.Y2015M10A.245>
- [28] Han-Joo Kim. 2017. *(PyeongChang 2018) Cutting-edge technologies to greet Winter Olympics visitors*. Retrieved September 11, 2021 from <https://en.yna.co.kr/view/AEN20170120006500320>
- [29] Sungmoon Kim. 2009. Trouble with Korean Confucianism: Scholar-Official Between Ideal and Reality. *Dao* 8, 1 (March 2009), 29–48. <https://doi.org/10.1007/s11712-009-9105-1>
- [30] Sang Yun Kim. 2019. The Fourth Industrial Revolution: Trends and Impacts on the World of Work. In *Handbook of Vocational Education and Training*, Simon McGrath, Martin Mulder, Joy Papier, and Rebecca Suart (Eds.). Springer International Publishing, Cham, 177–194. [https://doi.org/10.1007/978-3-319-94532-3\\_115](https://doi.org/10.1007/978-3-319-94532-3_115)
- [31] Yoochul Kim. 2018. *What it felt like to visit the most tech-centric Olympics ever*. Retrieved September 11, 2021 from <https://www.technologyreview.com/2018/02/23/104711/what-it-felt-like-to-visit-the-most-tech-centric-olympics-ever/>
- [32] Haksoo Ko and Joanne Koo. 2010. Cultural Values and the Korean Negotiator. *Journal of Korean Law* 9, 2 (June 2010), 225–244. <https://s-space.snu.ac.kr/>

- bitstream/10371/85169/1/03H\_Ko\_and\_J\_Koo.pdf
- [33] Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. Accountable Algorithms. *University of Pennsylvania Law Review* 165 (2017). [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2765268](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765268)
- [34] Ohbyung Kwon, Kyungwha Hwang, and Yun Seon Kim. 2019. Five Ethical Issues of the Fourth Industrial Revolution. *Logos Management Review* 17, 3 (Sept. 2019), 31–48. <https://doi.org/10.22724/LMR.2019.17.3>
- [35] Hyo Jin Lee. 2021. *Controversial chatbot leaves lessons on AI use ethics*. Retrieved October 19, 2021 from [https://www.koreatimes.co.kr/www/tech/2021/10/133\\_302390.html](https://www.koreatimes.co.kr/www/tech/2021/10/133_302390.html)
- [36] Jongho Lee and Keun Lee. 2021. Is the fourth industrial revolution a continuation of the third industrial revolution or something new under the sun? Analyzing technological regimes using US patent data. *Industrial and Corporate Change* 30, 1 (June 2021). <https://doi.org/10.1093/icc/dtaa059>
- [37] Sora Lee and Woojin Kang. 2021. Precision Regulation Approach: A COVID-19 Triggered Regulatory Drive in South Korea. *Frontiers in Public Health* 9 (Feb. 2021). <https://doi.org/10.3389/fpubh.2021.628073>
- [38] Sora Lee and Ryan Wong. 2021. COVID-19 Responses of South Korea as Hybrids of Governance Modes. *Frontiers in Public Health* 9 (Sept. 2021). <https://doi.org/10.3389/fpubh.2021.65495>
- [39] H.Y. Li, J.T. An, Y. Zhang, K.H.M. Mansur, and Y. Fu. 2021. Ethical Problems and Countermeasures of Artificial Intelligence Technology. *E3S Web of Conferences* 251 (2021). <https://doi.org/10.1051/e3sconf/202125101063>
- [40] Arend Lijphart. 1971. Comparative Politics and the Comparative Method. *The American Political Science Review* 65, 3 (1971), 682–693. <https://doi.org/10.2307/1955513>
- [41] Intralink Limited. 2018. Internet of Things South Korea Market Intelligence Report. Report. Retrieved September 11, 2021 from <https://www.intralinkgroup.com/getmedia/d06c063d-738b-4a05-8c8b-1c7fe379fbb4/Korean-Internet-of-Things-Final-Report,-Technology,-Brochure>
- [42] Simon Matti. 2009. *Exploring public policy legitimacy: a study of belief-system correspondence in Swedish environmental policy*. Ph.D. Dissertation. Luleå University of Technology, Luleå, Sweden.
- [43] Justin McCurry. 2021. *South Korean AI chatbot pulled from Facebook after hate speech towards minorities*. Retrieved September 22, 2021 from <https://www.theguardian.com/world/2021/jan/14/time-to-properly-socialise-hate-speech-ai-chatbot-pulled-from-facebook>
- [44] Katharina Miller and Karen Wendt. 2021. *The Fourth Industrial Revolution and Its Impact on Ethics: Solving the Challenges of the Agenda 2030* (1st. ed.). Springer International Publishing, Cham.
- [45] Erik Mobernd. 2019. *Top-Down Democracy in South Korea*. University of Washington Press, Seattle, WA.
- [46] Mark H. Moore. 1995. *Creating Public Value: Strategic Management in Government* (revised ed.). Harvard University Press, Cambridge, MA.
- [47] Jessica Morley, Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander, and Luciano Floridi. 2021. Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds and Machines* 31, 2 (June 2021), 239–256. <https://doi.org/10.1007/s11023-021-09563-w>
- [48] Jessica Morley, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Sci Eng Ethics* 26 (2020), 2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>
- [49] Jessica Morley, Libby Kinsey, Anat Elhalal, Francesca Garcia, Marta Ziosi, and Luciano Floridi. 2021. Operationalising AI ethics: barriers, enablers and next steps. *AI & Soc* 26 (2021). <https://doi.org/10.1007/s00146-021-01308-8>
- [50] Wiebe Nauta, Jaekwang Han, and Taekyoon Kim. 2021. Inspiring Democratic Progress in Development Assistance: South Korea's Aid Policy Reforms via Civic Engagement. *Forum for Development Studies* 48, 2 (May 2021), 309–330. <https://doi.org/10.1080/08039410.2021.1907784>
- [51] OECD. 2017. *PRESIDENTIAL COMMITTEE ON THE FOURTH INDUSTRIAL REVOLUTION | STIP Compass*. Retrieved September 11, 2021 from <https://stip.oecd.org/stip/policy-initiatives/2019%2Fdata%2FpolicyInitiatives%2F16688>
- [52] OECD. 2019. *The OECD AI Principles*. Retrieved September 11, 2021 from <https://www.oecd.org/digital/artificial-intelligence/ai-principles/>
- [53] OECD. 2019. Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO). *OECD Digital Economy Papers* 291 (Nov. 2019). <https://doi.org/10.1787/d62f618a-en>
- [54] Ministry of Foreign Affairs of the Republic of Korea. 2018. Korea PyeongChang Night” held during World Economic Forum Annual Meeting Mesmerizes People around World. Retrieved September 11, 2021 from [https://www.mofa.go.kr/eng/brd/m\\_5676/view.do?seq=319661](https://www.mofa.go.kr/eng/brd/m_5676/view.do?seq=319661)
- [55] National Museum of Korea. 2021. *Samgangaengsildo, An Illustrated Guide to the Three Bonds*. Retrieved November 30, 2021 from <https://www.museum.go.kr/site/eng/relic/represent/view?relicId=2831>
- [56] Personal Information Protection Commission of Korea. 2021. AI Personal Information Protection Self-Checklist (for Developer and User Use). Document.
- [57] Ministry of Science & ICT of Korea. 2020. Human-Centered AI Ethics Guidelines. Document. Retrieved March 23, 2021 from <https://www.4th-ir.go.kr/article/detail/1197?boardName=internalData&category=agenda>
- [58] Ministry of Science & ICT of Korea. 2020. MSIT releases human-centered 'National Guidelines for AI Ethics' draft. Report.
- [59] Ministry of Science & ICT of Korea. 2021. Plan for Implementation of Human-Centered, Trustworthy AI. Report. Retrieved September 27, 2021 from [https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=112&pageIndex=&bbsSeqNo=94&nttSeqNo=3180239&searchOpt=ALL&searchTxt=](https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=112&pageIndex=&bbsSeqNo=94&nttSeqNo=3180239&searchOpt=ALL&searchTxt=&pageIndex=&bbsSeqNo=94&nttSeqNo=3180239&searchOpt=ALL&searchTxt=)
- [60] Ministry of Science & ICT of Korea. 2021. Policy Seminar for Implementation of Human-Centered AI. Report. Retrieved September 27, 2021 from [https://policy.nl.go.kr/search/searchDetail.do?rec\\_key=SH2\\_PL20210264305&kwd=](https://policy.nl.go.kr/search/searchDetail.do?rec_key=SH2_PL20210264305&kwd=)
- [61] Ministry of Science & ICT of Korea. 2021. Policy Seminar for Implementation of Human-Centered AI. Video. Retrieved September 27, 2021 from <https://www.youtube.com/watch?v=8c-vc4duXLo&t=565s>
- [62] Ministry of Science & ICT of Korea and Korea Information Society Development Institute (KISDI). 2020. National Guidelines for Artificial Intelligence (AI) Ethics. Document.
- [63] Ministry of Science & ICT of Korea and Korea's Presidential Committee on the Fourth Industrial Revolution. 2020. 'National Guidelines for AI Ethics' Draft Public Hearing. Video. Retrieved September 22, 2021 from <https://www.youtube.com/watch?v=TuiJKzYkkm>
- [64] The Government of the Republic of Korea. 2019. Toward AI World Leader beyond AI: National Strategy for Artificial Intelligence. Report. Retrieved September 12, 2021 from <https://english.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=10&mPid=9&bbsSeqNo=46&nttSeqNo=9> Government Report No. 11-1721000-000393-01.
- [65] Seong-Mok Oh. 2018. *KT showcases 5G innovation at the Olympics in PyeongChang*. Retrieved September 11, 2021 from <https://news.itu.int/kt-showcase-5g-olympics/>
- [66] Chisung Park and Jooha Lee. 2020. Stakeholder framing, communicative interaction, and policy legitimacy: anti-smoking policy in South Korea. *Policy Sciences* 53, 4 (Dec. 2020), 637–665. <https://doi.org/10.1007/s11077-020-09394-z>
- [67] Chisung Park, Jooha Lee, and Changho Chung. 2015. Is “legitimized” policy always successful? Policy legitimacy and cultural policy in Korea. *Policy Sciences* 48, 3 (Sept. 2015), 319–338. <https://doi.org/10.1007/s11077-015-9220-2>
- [68] Kyung Ryul Park, Sundeep Sahay, Jørn Braa, and Pamod Amarakoon. 2021. Digital Resilience for What? Case Study of South Korea. In *Proceedings of the 1st Virtual Conference on Implications of Information and Digital Technologies for Development (1st Virtual Conference on Implications of Information and Digital Technologies for Development)*. Working Group 9.4 of the International Federation of Information Processing (IFIP), 139–152. <https://arxiv.org/ftp/arxiv/papers/2108/2108.09950.pdf>
- [69] Huw Roberts, Josh Cows, Jessica Morley, Mariarosaria Taddeo, Vincent Wang, and Luciano Floridi. 2021. The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & Soc* 36 (March 2021), 59–77. <https://doi.org/10.1007/s00146-020-00992-2>
- [70] Francesca Rossi and Nicholas Mattei. 2019. Building Ethically Bounded AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. AAAI, 9785–9789. <https://doi.org/10.1609/aaai.v33i01.33019785>
- [71] Gilbert Rozman. 2002. Can Confucianism Survive in an Age of Universalism and Globalization? *Pacific Affairs* 75, 1 (2002), 11–37. <https://doi.org/10.2307/4127239>
- [72] Oscar Schwartz. 2019. *In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation*. Retrieved September 22, 2021 from <https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation>
- [73] Gerry Stoker. 2006. Public Value Management: A New Narrative for Networked Governance? *The American Review of Public Administration* 36, 1 (March 2006), 41–57. <https://doi.org/10.1177/0275074005282583>
- [74] Deborah Stone. 2002. *Policy Paradox: The Art of Political Decision Making* (3rd. ed.). Norton & Company, New York, NY.
- [75] Tae Kyung Sung. 2018. Industry 4.0: A Korea perspective. *Technological Forecasting and Social Change* 132 (July 2018), 40–45. <https://doi.org/10.1016/j.techfore.2017.11.005>
- [76] Hamed Taherdoost. 2016. Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *SSRN Electronic Journal* (2016). <https://doi.org/10.2139/ssrn.3205035>
- [77] Jacob Torfing, B. Guy Peters, Jon Pierre, and Eva Sørensen. 2012. *Interactive Governance: Advancing the Paradigm* (3rd. ed.). Oxford University Press, Oxford.
- [78] Weiming Tu. 1993. Probing the 'Three Bonds' and 'Five Relationships'. In *Confucianism and the Family: A Study of Indo-Tibetan Scholasticism*. Walter H. Slote and George A. De Vos (Eds.). SUNY Press, Albany, NY.
- [79] Jess Whittlestone, Rune Nyrup, Anna Alexandrova, and Stephen Cave. 2019. The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AI/ES '19)*. Association for Computing Machinery, New York, NY, USA, 195–200. <https://doi.org/10.1145/3306618.3314289>

## A LIST OF INTERVIEWEES

By gathering qualitative data in the form of 9 interviews, I could observe policy stakeholders' roles and subjective interpretations concerning not only the Guidelines' development process, but also the Guidelines' societal significance and impact (see Table 1). According to [67], these elements are important to policy legitimacy (see Section 2.2). I was able to gather this group of interviewees via snowball and judgment sampling [20, 76].

Whereas one interview (Ministry of Science and ICT director) was conducted in-person, the others were conducted online via videoconferencing platform Zoom, and each interview lasted between 40 minutes and one hour and 10 minutes. One interview was conducted mostly in English (computer scientist) while the remaining interviews were conducted in Korean. All online interviews were video-recorded using Zoom's recording feature. These interviews were semi-structured and open-ended, wherein the questions, while initially standardized, could be reformulated according to interview flow and the answers were open ended.

**Table 1: List of Stakeholder Interviewees**

No.	Interviewee	Position	Affiliation	Interview Date
1	A	Professor	Korea Advanced Institute of Science & Technology (KAIST), School of Computing	May 11, 2021
2	B	Director of the Artificial Intelligence Policy Bureau	Ministry of Science & ICT of Korea	June 10, 2021
3	C	Director	Institute for Digital Rights	July 2, 2021
4	D	Director	Korea Information Society Development	July 7, 2021
5	E	Professor, President	Kyungnam University, The Korea Association for Posthuman Society	July 20, 2021
6	F	Professor	Seoul National University, Law Department	July 20, 2021
7	G	Director	Artificial Intelligence Industry Association of Korea	August 6, 2021
8	H	Professor	Kwangwoon University, College of Public Policy & Law - Division of Law	August 10, 2021
9	I	Professor	Seoul National University of Education and Ethics Education	August 19, 2021

## B GOVERNMENT DOCUMENTS

**Table 2: Summary of publications referred to for South Korea as case study**

Official Government Publications	Publicly Available Discussions, Hearings and Seminars
Toward AI World Leader beyond AI: National Strategy for Artificial Intelligence	'National Guidelines for AI Ethics' Draft Public Hearing (국가인공지능윤리기준(안) 공청회)
National Guidelines for Artificial Intelligence (AI) Ethics (국가 인공지능 윤리기준)	Policy Seminar for Implementation of Human-Centered AI (사람중심의인공지능(AI) 구현을위한 정책세미나)
Human-Centered AI Ethics Guidelines (사람이중심이되는 인공지능(AI) 윤리기준)	Legislative Hearing for Trust and for Act on Creating Foundation Fostering Artificial Intelligence (인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안 입법공청회)
MSIT releases human-centered 'National Guidelines for AI Ethics' draft.	
Plan for Implementation of Human-Centered, Trustworthy AI (사람이 중심이 되는 인공지능을 위한 신뢰할 수 있는 인공지능 실현 전략(안))	
AI Personal Information Protection Self-Checklist (for Developer and User Use) (인공지능(AI) 개인정보보호자 점검검표(개발자, 운영자용))	
Basic Principles for Protection of Users of AI-Based Media Recommendation Services (인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙)	