Join us on **Facebook**
Follow us on **Twitter**

Related Links

CHINA
AI ambitions revealed by most cited research papers
ASIA
AlphaGo computer victory sparks AI 'deep learning' drive

Featured Employers

جامعة الإمارات العربية المتحدة
United Arab Emirates University

ASIA

Robots bring Asia into the AI research ethics debate

Yojana Sharma 24 November 2017

Universities in China and elsewhere in Asia are belatedly joining global alliances to promote ethical practices in artificial intelligence or AI, which were previously being studied in university research centres in a fragmented way.

Countries like South Korea, Japan, China and Singapore are making huge investments in AI research and development, including the AI interface with robotics and are in some areas rapidly narrowing the gap with the United States. But crucially there are still no international guidelines and standards in place for ethical research, design and use of AI and automated systems.

China's universities in particular are turning out a large number of researchers specialising in AI. Whereas in the past they would head for Silicon Valley in the US, many are now opting to stay in the country to work for home-grown technology giants such as Alibaba, Tencent and Baidu – companies which gather and use huge amounts of consumer data with few legal limits.

In July Chinese leader Xi Jinping unveiled a national plan to build AI into a US$152.5 billion industry by 2030 and said the country was aiming for global dominance.

"China's pace of AI research and adoption is astoundingly fast, it is perhaps the market that adopts AI technology the quickest, so there is a lot of advanced research being done," Pascale Fung, a professor in the department of electronic and computer engineering at Hong Kong University of Science and Technology, or HKUST, told *University World News*.

"Our prime concern is to look at the ethical adoption of AI in terms of setting up standards. Do we also need regulations; if so, what? This conversation has not happened in this region yet.

"There is no transparency about dataflow. And there is no certification of AI safety," she says.

Major US technology companies Google, Facebook, Amazon, IBM and Microsoft last year set up an industry-led non-profit consortium **'Partnership on AI to Benefit People and Society'** to come up with ethical standards for researchers in AI in cooperation with academics and specialists in policy and ethics.

HKUST announced earlier this month it had become the first Asian university partner in the alliance. The previous lack of Asian participation, academic or otherwise, is surprising considering the fast pace of AI developments in the region.

The global focus on AI ethics is "only starting and it is an international effort but with very little participation from Asian countries", says Fung. "My role is to bring the top adopters of AI technology, namely the East Asian countries, to the table and to co-lead this effort."

Liability and legal entity

Researchers have also become concerned about regional efforts, such as in the European Union, to regulate AI systems, particularly those driving robots, to establish liability. The European Parliament, for example, has put forward **ideas to recognise robots as a legal entity**, such as in the case of driverless cars.

It was announced last week that a robot developed by a leading Chinese AI company, iFlytek, passed the written test of China's national medical licensing examination. Although iFlytek said its robot is not intended to replace doctors but to assist them, it has brought the

issue of AI ethics to the fore in a country with a massive shortage of doctors, particularly in rural areas.

“We advocate that AI should not be the one making life-and-death decisions. AI can advise the medical doctors who in turn are the ones certified to practise medicine,” Fung says. “But so far these ideas have not yet been adopted internationally.”

“There need to be good practice guidelines and standards of how we use AI, for example in healthcare. Right now there are absolutely no guidelines. We are just playing it by ear,” says Fung. “If we don’t start working on this now, I am afraid there will be a huge accident and then the regulations will come and that will be a bit too late.”

The World Economic Forum’s ***Global Risks Report 2017***, which surveyed 745 leaders in business, government, academia, non-governmental and international organisations, including members of the Institute of Risk Management, named AI and robotics as "the emerging technology with the greatest potential for negative consequences over the coming decade".

Work on global standards

Fung believes Asian involvement in setting ethical guidelines is essential if globally acceptable guidelines are to be adopted within the region. “There are standards associations around the world and they are international, but that has been very little participation so far by East Asian countries, including China,” she notes.

The main work on global standards and ethical best practice for automated and intelligent systems is being carried out by the Institute of Electrical and Electronics Engineers or IEEE.

“Our ambition is to make it possible for the technical and scientific community to take into account at least the values of society and right now this is not done,” says Konstantinos Karachalios, managing director at IEEE Standards Association.

The race to be first in developing AI systems “is the big temptation of our time, just do it before others do it”, Karachalios adds. The assumption is that what is being researched and developed is good and the prevailing view is “if there is a problem with the final project it is not our problem, it is the damn people who use it”, Karachalios told *University World News*. “This is wrong.”

The first version of the **IEEE’s global standards** released last year incorporated the views of more than 150 experts in AI, law, ethics and policy. But it was seen as based largely on Western principles. This is being rectified with a new version to be released next month based on feedback, including from non-Western countries, particularly in Asia.

Cultural sensitivity is key for universal adoption of ethical standards into the design of systems. Karachalios says the need is for ethical standards to be incorporated, "but we don’t say which values to embed”.

Sara Mattingly Jordan, an assistant professor in the Center for Public Administration and Policy at Virginia Tech in the US, who is collating the inputs and responses to the IEEE standards document, says AI ethics “is still very much an intellectual’s topic”, involving mainly university academics.

Within the AI industry, “right now we are relying on people’s professional judgment and professional expertise at an individual level of ethics. That’s what’s controlling the system right now and it’s pretty fragile.”

“The hazard of people working in teeny tiny disaggregated teams with global reach in a vacuum is a serious potential hazard,” she says. “If each individual nation or each individual university tries to publish its own code of ethical data standards, how is anybody going to operate as a vendor in that system? It’ll create substantial problems.”

Including China

But companies, including law firms, are now beginning to join the debate and the need to include the major Asian AI powerhouses – South Korea, Japan, Singapore and China – is also recognised. “It would be great if we can get China on board; nobody disputes that they are a major player,” she says. “But that doesn’t mean that we are demanding that China change its perspective.”

Experts say the Chinese government would balk at any legalistic rules or guidelines that question the supremacy of the state to control such technologies, as well as anything that smacks of individual privacy rights that might supersede the right of government over its citizens.

“There is a big interest from their [the Chinese] side to engage with the ethical aspects and not the political,” says the IEEE’s Karachalios. “The political dimension is involved because in the end it is about freedom and freedom also has an ethical dimension. This may not be something that is interesting for them and we must respect it.

Top Stories Last Week

EUROPE
Research assessment reform efforts poised for breakthrough

GLOBAL
Updated Magna Charta marks

"We must still find a way to engage with each other and have a fruitful dialogue," he says, and points out that "our standards are not laws, they are recommendations from peer to peer".

If producers of AI systems "can show that they can produce a product that is trustworthy and respects privacy then maybe people will preferentially choose it, even if they make it more expensive because they use more time and energy looking at these [ethical] aspects," Karachalios says.

With its global AI ambitions, China definitely wants to be part of the process, says HKUST's Fung. "On standards and regulations you can bet the Chinese don't want to be left out."

Receive UWN's free weekly e-newsletters

Please fill in your details below to subscribe to our service.

Email
Title Select...
First Name
Last Name
Job Title Select ...
Organisation
Country Select a country ...

☐ I check my email on my phone

Please send me the following emails:

☐ Global Newsletter (weekly)

☐ Africa Newsletter (weekly)

☐ **UWN event invitations and other UWN emails**
To be able to support its journalism, UWN relies on being able to invite readers to participate in its webinars and other events supported by partnerships.

☐ I agree that my data will be stored and processed according to UWN's terms & conditions.

Please click the reCAPTCHA checkbox below

I'm not a robot reCAPTCHA Privacy - Terms

Submit Subscription

shift in universities' thinking

UNITED STATES-CHINA-AFRICA
Recruiting in Africa: US faces a stiff competitor in China

MOROCCO
Nearly half of respondents in study admit to plagiarism

GLOBAL
Universities should make democracy education a requirement

GLOBAL
Science diplomacy's key role building peace: DAAD chief

IRAN
Campuses erupt over death of woman held for 'loose hijab'

UNITED STATES
Gates' gift to HBCUs: A $100m investment in diversity

KENYA
Burnout cases on the increase in academia, says report

ZIMBABWE
Zim students ask court to suspend hike in tuition fees

Perspectives and Approaches in AI Ethics: East Asia

Danit Gal

The Oxford Handbook of Ethics of AI
Edited by Markus D. Dubber, Frank Pasquale, and Sunit Das

Print Publication Date: Jul 2020 Subject: Law, IT and Communications Law
Online Publication Date: Jul 2020 DOI: 10.1093/oxfordhb/9780190067397.013.39

Abstract and Keywords

This chapter focuses on Chinese, Japanese, and South Korean perspectives on and approaches to AI and robots, which can be tools and partners in the AI ethics debate. Each country, in its own way, debates its movement across the tool-partner spectrum. To date, South Korean policy makes a stand against partner AI and robots, while popular culture explores the idea. Chinese policy is headed in the direction of a tool-oriented AI and robotics ethical guidelines, while local practices and culture experiment with the idea of physical and spiritual partnership. Meanwhile, Japan's social principles are also moving in the tool direction, but its society actively seeks and creates partner-like AI and robots. The chapter then considers two cross-cutting AI and robotics-related ethical issues: the Anthropomorphized Tools Paradox and female objectification. These issues underscore the question of "antisocial" technology. It is clear that both the Anthropomorphized Tools Paradox and female objectification in technology fall under "antisocial" development.

Keywords: China, Japan, South Korea, AI ethics, partner AI, robotics ethics, Anthropomorphized Tools Paradox, female objectification, antisocial technology

Introduction

FOR centuries humanity has been building technological tools to support and enhance its capabilities, allowing us to survive and flourish. For years humanity has dreamed about creating others in its image, leading to a history rich with human statues, automata, robots, and artificial intelligence (AI). As we usher in another era of technological development, we are faced with the social consequences of our dreams—a technology that exceeds the status of tool and is moving toward that of a partner. Let us be clear: AI and robots themselves are tools. And yet their perception is increasingly that of partners, blurring the line between what they are and what they could be. Partnership is a broad term in this context. It encompasses tools functioning as social caretakers, friends, companions, romantic love interests, and fellow spiritual beings. All of which, developed and used in East Asia, are briefly surveyed in this introductory chapter.

The framing of this chapter is founded on the claim that China, Japan, and South Korea perceive and approach AI and robots on a spectrum ranging from tool to partner. This may be true for other countries, but is distinct in East Asia. While continuously moving on this spectrum, the policy, academic thought, local practices, and popular culture observed in each country place South Korea in the tool range, China in the middle, and Japan further along in the partner range.

The tool perspectives on and approaches to AI and robotics are seemingly common in the West, where the technology is viewed as an instrument for development and (p. 608) growth. This is also true for East Asia, especially where official government and corporate policies are concerned. The divergence toward the partner range occurs in academic thought, local practices, and popular culture.

Some readers may be tempted to categorize the partner range of the spectrum as a simple act of anthropomorphism, defined as "to attribute human form or personality to things not human,"[1] but that is an oversimplification. While it is true that East Asian countries demonstrate a preference toward biologically inspired (humanoid and animaloid) AI and robots, they do much more than that. Under Buddhist and Shinto techno-animism, AI and robots are not just attributed human traits, they are believed to possess a spirit or spiritual essence. In addition, in East Asia, AI and robot partners have come to be viewed as friends and love interests. These spiritual, psychological, and emotional perceptions of AI and robots, in turn, are further amplified by the fact that AI and robots are designed to not only look but also behave like us.

The human tendency to anthropomorphize runs into further complications when AI and robots are designed to have the relatable interface of a potential partner but are intended be used as a tool. This creates a functional and emotional paradox where designing tools mimicking humans is desirable, but developing natural human emotions in response to such mimicking is seen as problematic. The chapter names this design contradiction the Anthropomorphized Tools Paradox. This paradox, often tightly knit with issues of female objectification, constitutes another notable source of growing sociotechnical tensions. These tensions, emanating from our movement on the tool-partner spectrum, have inspired numerous global AI and robotics ethics debates throughout the years. And yet in no place are these debates as feasible and as socially pressing as in East Asia.

South Korea

South Korea is placed in the tool range due 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. A caveat in this hierarchy is that within the human layer exist additional social hierarchies that may compromise the inclusive potential of AI and robots. South Korea also demonstrates a clear preference for functional AI applications and robots, mainly focusing on human empowerment in the areas of public services like education, healthcare, social care, disaster relief,

and security. A divergence toward the partner range is found in the country's popular culture, which is awash with human-AI-robots partnership stories.

(p. 609) Policies and Ethical Principles: Tool-Decisive

In 2007, the Korea Institute for Robot Industry Advancement (KIRIA) published a Robots Ethics Charter, which was further revised in 2016.[2] These past iterations served as the foundation of an all-government ethical framework adopted by the South Korean National Information Society Agency (NIA). The NIA released its Ethics Guideline for the Intelligent Information Society in April 2018.[3] The guideline aims to achieve a human-oriented intelligent information society through the PACT Principles: Publicness, Accountability, Controllability, and Transparency.

Building on the 2007 Robot Ethics Charter's comprehensive ethical guidelines for developers, providers, and users, the Ethics Guideline assumes four unique positions: (1) It places considerable responsibility on users, calling for informed and self-regulated use. (2) It places the responsibility for preemptively assessing AI and robots' potential negative social impact on providers. (3) It holds developers responsible for the elimination of socially discriminatory characteristics in AI and robotics design and for making it accessible to disadvantaged and vulnerable groups. Developers are also placed in the vanguard of AI ethics in South Korea. (4) It calls to refrain from developing AI and robots with "antisocial" characteristics and to "minimize social resistance and disorder against the universal use" of AI and robots.[4]

In addition, the Ethics Charter published by the NIA in June 2018[5] reiterates the six principles codified in the 2016 version of the Robots Ethics Charter by KIRIA. Notable among these principles is the balance between the protection of human dignity (first principle) and the common good (second principle). Sunyoung Byun, professor at the Seoul National University of Education, explains that the three versions of the Ethics Charter approach AI and robots as tools meant to protect human dignity and promote the common social good. He notes the difficulty humans have in maintaining a harmonious balance between individual and collective flourishing and that this balancing act constitutes an important moral dilemma further complicated by the introduction of AI and robots.[6] These documents spell a clear human-over-machine hierarchy where humans (developers, providers, and users) protect each other and AI and robots protect and service humans.

(p. 610) Academic Thought and Local Practices: Tool-Oriented

South Korea's tool perspectives on and approaches to AI and robots is also evident in its 2018 Winter Olympics technology demonstration, where it deployed eighty-five functional robots with varying degrees of intelligence. Rescue robots from the Korea Advanced Institute of Science and Technology (KAIST) were used as torchbearers alongside celebrities, and robots serviced the airport and different competition venues.[7] Functional robots are not immune to ethical clashes and debates, as seen in the international boycott of KAIST by over fifty AI researchers due to its partnership with a defense company. The ban was

lifted shortly after KAIST ensured the researchers it will not develop or assist in the development of lethal autonomous weapon systems or killer robots.[8] Two months later, KAIST announced the establishment of an AI ethics subcommittee.[9] But Chi Hyung Jeon, assistant professor at KAIST, states that the subcommittee was formed during the time of the boycott in April 2019 and hosted its first event with the boycott organizers' participation, drawing a line between the two events.[10]

Further in conjunction with the ban, KAIST released a Code of Ethics for Artificial Intelligence on April 4, 2018.[11] Unique among its four principles is the third one, stipulating that "AI shall follow both explicit and implicit human intention. However, before execution, the AI should ask people to confirm the implicit intention. (If several people are involved and their intentions are different, the AI should follow the person with the highest priority or closest relationship.)" This suggests additional human hierarchies within the human layer of the human-over-machine hierarchy, where "highest priority" people will have the final say on navigating the AI as they see fit. KAIST's principle further complicates an established issue of societal inequality in South Korea[12] by reinforcing positions of power or relationship which conflict with other documents calling for more equally distributed and accessible AI and robots. In particular, given KAIST's position as an educational institute, this principle may also conflict with developers' mandate to act as eliminators of social bias and discrimination under NIA's Ethical Guideline.

(p. 611) Popular Culture: Between Partnership Exploration and Tool Preference

But while the KAIST code of ethics may challenge NIA's desired harmony within the human layer of the hierarchy, South Korean popular culture is challenging the hierarchy itself. Korean dramas offer a more controversial perspective by entertaining the idea of AI and robots as being more than just tools. The country saw eight dramas about AI and robots acting as family members, friends, and love interests between 2016 and 2019. This list includes: 사랑하면 죽는 여자 봉순이 or Bong Soon—a Cyborg in Love (2016); 아이엠 or I am … (2017); 로봇이 아니야 or I'm Not a Robot (2017); 보그맘 or Borg Mom (2017); 109 별일 다 있네 or 109 Strange Things (2017); 너도 인간이니 or Are you Human Too? (2018); 사랑은 사람처럼 or Love Like a Person (2019); and 절대 그이 or Absolute Boyfriend (2019), an adaptation of a Japanese story. While most shows conclude that human companionship is superior to that of AI and robots, they frequently and publicly explore the idea of AI and robots transitioning from tools to partners.

Directly addressing this disruption to the human-over-machine hierarchy, Jiwon Kim, head of AI Policy at the Ministry of Science and ICT, notes that "as people become more reliant on and overuse intelligent social robots, the risk of losing the basic ethical values we hold as humans, as well as authentic human relationships, increases." Kim states that they therefore "believe that attachment to obedient robots could undermine people's relationships with other humans."[13] This concern may be key in explaining the NIA's emphasis on avoiding the "antisocial" development of AI and robots. This raises an important question: what degree of AI and robotics development is perceived as "antisocial" by virtue of its

ability to disrupt human relationships? This question will become increasingly pressing as we examine Chinese and Japanese perspectives and approaches.

Such concerns may also be a response to the findings of a 2018 study on human interactions with a social intelligent robot in South Korea.[14] But most South Koreans, it seems, still prefer functional AI and robots. A 2016 study found that South Korean respondents preferred a functional robot over a biologically inspired one when compared with their Japanese peers. Respondents maintained that a functional robot made them feel like they had more control over it. This sense of hierarchical preservation is believed to have induced an increased sense of comfort among South Korean users.[15] The observed sense of comfort aligns with the NIA's wishes to "minimize social resistance (p. 612) and disorder against the universal use" of AI and robots.[16] But as more systems take on humanoid and animaloid shape to make their interface relatable and master biologically inspired capabilities to better communicate with humans, this sense of control and comfort will likely erode.

Conclusions

This chapter places South Korea in the tool range of the spectrum quite comfortably. And yet it does note that the established human-over-machine hierarchy will face continued challenges as the technology evolves and the human imagination continues to run wild. In fact, the adherence to a human-over-machine hierarchy and the calls to avoid "antisocial" development further highlight the debate on human-AI-robots partnerships. It also highlights the fact that further social hierarchies exist within the human layer of said human-over-machine hierarchy. This amplifies existing social tensions that are likely to increase in complexity and importance as more artificial actors join the societal mix. A human-machine integration is inevitable given South Korea's plans to create an intelligent information society, even if it's a human-oriented one. Policymakers in South Korea and elsewhere should thus remember that technology does not solve social problems, it typically further exacerbates them.

China

Much like South Korea, China holds a top-down view of AI and robots as tools for progress, as demonstrated in official government and corporate policies and recommendations. The ethical components of this approach are beginning to materialize, and evidence suggests that they align with other global guidelines, largely viewing AI and robots as tools. China, however, also presents a strong interest in imbuing AI and robots with partner-like capabilities to help them realize their full positive potential. This is apparent in its academic thought, local practices, and popular culture. As such, China is also home to intensifying tensions between top-down tool and bottom-up partner approaches and perspectives that will shape and inform future local debates and practices.

Policies and Ethical Principles: Tool-Oriented

Published on July 20, 2017, the Chinese Government's New Generation AI Development Plan calls for the establishment of ethical norms and frameworks. Namely, it calls for the establishment of an "ethical and moral multi-level judgement structure and (p. 613) human-computer collaboration ethical framework" and "an ethical code of conduct."[17] According to the plan's timeline, the codification of these ethical norms, codes, and frameworks is slated to take place between 2025 and 2030, during the last stretch of the development plan.

This will not take as long. On January 10, 2019, the Chinese Association for Artificial Intelligence (CAAI), the only state-level AI organization sitting under the Ministry of Civil Affairs,[18] announced the establishment of an AI ethics committee tasked with creating guidelines for Chinese development.[19] The committee is led by Professor Xiaoping Chen, known for leading the creation of a realistic female humanoid "robot goddess" named Jia-Jia (which embodies both the Anthropomorphized Tools paradox and female objectification issues).[20] Chen explained the uniqueness in AI and robot ethics, saying that "a smart humanoid robot could integrate into people's daily lives someday, but no one knows for sure what kind of risks it may bring along with its service."[21] This suggests a challenging ethical balancing act between technology created to act as an intelligent tool but designed with the characteristics of a desirable partner.

In early March 2019, the Chinese government hosted the 13th National Committee of the Chinese People's Political Consultative Conference and the 13th National People's Congress, known as the Two Sessions. During the Two Sessions, the CEOs of Baidu and Tencent submitted proposals discussing AI ethics.[22] Baidu's Robin Li Yanhong submitted a proposal calling on the government to speed up AI ethics research, citing the impending transition of AI from a simple tool to a stakeholder in many areas. In particular, Li urged the government to share distinct Chinese wisdom with the international AI ethics community.[23] On May 26, 2018, Li also introduced Baidu's four AI ethics principles: safety and controllability, equal access, human development, and freedom.[24] Tencent's Pony Ma Huateng submitted a proposal calling for ethical AI regulations and (p. 614) the use of AI for social good.[25] This stance is in line with the Tencent Research Institute's (TRI) running project: AI for Social Good.[26] The TRI also published an AI ethics framework for multiple stakeholders under the ARCC principles: Available, Reliable, Comprehensive, and Controllable.[27]

On May 25, 2019, a group of leading Chinese institutions including the Beijing Academy of AI, Peking and Tsinghua universities, the Chinese Academy of Science, and industry leaders like Baidu, Tencent, and Alibaba released the Beijing AI Principles.[28] Aligning with other existing AI principles, the Beijing Principles emphasize AI development to benefit humanity. A notable suggestion in the principles maintains that "stakeholders of AI systems should be able to receive education and training to help them adapt to the impact of AI development in psychological, emotional, technical aspects."[29] This suggests a considerable degree of expected psychological and emotional interactions between hu-

mans the AI systems, rather than just functional ones. A statement of this sort indeed hovers between the tool and partner range of the spectrum.

Considering the observed alignment with international AI ethics discourse, what would the aforementioned distinct Chinese wisdom to be brought into these discussions look like? Miao Liao, lecturer at Changsha University of Science and Technology, believes the answer is found in the pluralistic integration of the Chinese government's twelve "core socialist values."[30] These values are divided into three groups. National values include prosperity, democracy, civility, and harmony. Social values include freedom, equality, justice, and the rule of law. Individual values include patriotism, dedication, integrity, and friendship. These twelve values, she explains, are already integrated into a nationally taught graduate Engineering Ethics (工程伦理) textbook. The textbook also highlights four unique Chinese characteristics in comparison with Western engineering ethical guidelines: responsibility precedes freedom, obligation precedes rights, the group precedes the individual, and harmony precedes conflict.[31] It is highly likely that state-adopted and state-approved AI and robotics ethical guidelines will incorporate these values or, at the very least, reflect their spirit as in the case of engineering.

(p. 615) Academic Thought and Local Practices: Uniquely Partnership-Oriented

An AI ethics initiative departing from the tool approach to AI is the Harmonious Artificial Intelligence Principles (HAIP), led by Yi Zeng, professor at the Chinese Academy of Science (CAS) Institute of Automation. The HAIP code of ethics promotes unique concepts like: (1) humanization to strengthen interactions between AI and humans; (2) empathy and altruism to ensure a harmonious human-AI society; (3) human empathy toward the AI; (4) privacy for AI, which humans should respect; (5) bias against the machine, where humans should not show bias where both AI and humans show similar risks; and (6) legal constraints on how humans treat AI to ensure a harmonious coexistence.[32] Zeng claims that the safest approach to develop AI and robots is to give them a sense of self (consciousness) so that they are able to empathize with humans. He believes the reciprocity between humans and AI and robots is key in achieving true harmony to ensure the technology remains beneficial as it continues to evolve.[33] This approach to AI ethics marks a clear shift toward the partner range of the spectrum and serves as a rare demonstration of what ethical principles aiming to achieve this partnership vision might look like.

Zeng is not alone in his belief that AI systems should ascend to a higher level of consciousness. Hanniman Huang, a veteran Chinese AI product manager, views AI and robots as a new species. He sees them as a carrier for the human exploration of self-limitation and the relationship between heaven and man. Huang believes that the technologies' unique advantages will fully manifest when humans move from using them as a substitute (tool), to having them as a part of their society (partner), to coexisting with them (human-AI-robots symbiosis, found on the far end of the partner range). According to Huang, this symbiosis will be achieved when developers possess critical competencies in humanistic

and spiritual realms of knowledge by practicing Buddhism. He specifically points toward the Buddha dharma, which symbolizes the natural law and harmony and thus guides ethical behavior.[34] Huang therefore takes human-AI-robots partnerships an additional step further to include both physical and spiritual connections.

Popular Culture: Partnership-embracing

This aligns with the Chinese Buddhist idiom 万物皆有灵，念起莲花开 roughly translating into "everything has a soul, if you believe in Buddha the lotus (upon which Buddha sits) will bloom." This suggests that everything can be cultivated toward enlightenment and become the Buddha. This principle originated in ancient Indian Buddhist (p. 616) scriptures that have been adopted and adapted throughout Asia, now deeply embedded in Chinese tradition among others in the region.[35]

An application found in the intersection between Buddhism and popular Chinese culture is the intelligent robot monk, 贤二 or Xian'er, roughly translating into "simple looking but virtuous." Xian'er was introduced in October 2015 by the Longquan Temple's information and technology center as a preprogrammed robot meant to help spread the message of Buddhism.[36] In 2018, Xian'er received a machine learning boost to engage with Buddhist scripters and its over one million social media followers on a deeper analytical and conversational level.[37]

Chinese popular culture also provides notable views of AI and robots as partners. Chinese dramas have been depicting AI and robots as love interests since 1996 with 机器人趣话 or Funny Robot Talk (1996). Later dramas include the movie 机器侠 or Metallic Attraction Kungfu Cyborg (2009), 机器男友 or Robot Boyfriend (2017), 我的真芯男友or My Robot Boyfriend (2017), 天降机器女仆 or Robot Maid from Heaven (2017), and 我的保姆手冊 or Hi, I'm Saori (2018). A popular AI in China is the social chatbot XiaoIce (小冰), created by Microsoft and operated by Tencent. As the world's most popular social chatbot, XiaoIce has 660 million online users who often perceive it as a friend and love interest. Modeled after a female teenager (raising issues of female objectification and the depiction of minors), XiaoIce is liked enough to be considered among China's top celebrities.[38]

AI and robots are also present in the Chinese music scene, where a group of AI idols named May Wei VIV (五月薇VIV) was created based on the looks, talents, and personality traits of the Chinese idol group SNH48.[39] The idea of AI replicas is intended to help celebrities engage with their fans and continue entertaining tirelessly. The technology was showcased in China's Spring Festival Gala, where four famous human hosts were joined by their AI replicas on stage.[40] This also extends to the creation and use of three (p. 617) holographic AI news anchors on China's Xinhua state news agency.[41] These applications demonstrate that even if AI and robots are used as tools to create more engaging and accessible entertainment, audiences are likely to engage with these replica idols as friends and partners, much like they would like to engage with the humans they are modeled after. This also extends to robots designed with attractive human traits, even if not idols, like JiaJia.

Conclusions

Despite the compelling counternarrative academic thought, local practices, and popular culture present on human-AI-robots partnership, the majority of AI and robot applications in China are still perceived as tools as the country rushes to develop and apply these technologies. AI and robots are used across national sectors from healthcare and education to public services and military uses, and are expected to be fully embedded in the country's near future operations. However, as the nation's appetite for the rapid and effective development of AI and robots increases, their view as partners rather than tools is likely to continue increasing in prominence.

While the Chinese government has not alluded to the perception of AI as more than a tool, Chinese companies are taking part in developing social AI and robots that are perceived as partners. In addition, some prominent academics and developers are expressing ambitions that further expand previous conceptions of AI and robots as partners. China is therefore hovering around the middle of the tool-partner spectrum, with current policy, AI and robotics ethical considerations, and applications pulling in the direction of tool, and developers' ambitions and popular culture pulling in the direction of partner. This tension is expected to grow in local and global prominence as China pursues a leadership position in AI and robotics. It is also expected to aggravate social tensions as digital natives grow up with the Anthropomorphized Tools Paradox and female objectification of AI and robots intended to be used as intelligent tools, but designed to look and behave like desirable partners.

Japan

Japan sits in the partner range of the spectrum due to its exceptionally strong mix of pro human-AI-robots partnership academic thought, local practices, and popular culture. While the Japanese policy approach to AI is moving toward the tool range like South (p. 618) Korea and China, the extent of its societal vision for coexistence and coevolution with AI and robots is distinct. Another distinct feature in Japan is its strong techno-animistic tradition, which has likely inspired the development of its favorable partnership attitude. This entails a more intertwined and complex analysis of Japan's perspectives on and approaches to AI and robots.

Policies and Ethical Principles: Tool-Leaning

Like South Korea and China, Japanese policy views AI as a tool. It also, however, seeks to integrate AI and robots into all aspects of society to create an environment where humans, AI, and robots can coexist and coevolve. Japan's 5th Science and Technology Basic Plan, released on January 22, 2016, introduces the idea of Society 5.0.[42] This is a vision of an AI and robot-enabled, convenient, and diverse society that responds to all human needs and can even anticipate and respond to such needs before they emerge.[43] While this is likely the most progressive social vision in AI and robotics planning to date, the

idea of being able to anticipate and respond to human needs before they emerge constitutes a potential ethical issue. This may lead to a push rather than pull culture where humans will not necessarily decide what needs they want fulfilled and how.

The aforementioned vision is further fleshed out by the Cabinet Office Council on the Social Principles of Human-centric AI, in their Social Principles of Human-centric AI draft document. In the document, the council sets a social framework to guide the creation of an AI-ready Japanese society.[44] To this end, the document calls for the redesigning of society in "all aspects including Japan's social system, industry structure, innovation system, governance, and its citizen's character."[45] The document also, however, warns about overdependence on AI and robots and emphasizes the need to maintain human dignity when using them as tools, much like the South Korean guideline. And yet the document still calls for an "AI-based human living environment"[46] and a "society premised on AI."[47] A notable aspect of this document is its view of AI and robotics as widespread social tools that necessitate the redesigning of Japan's social systems and even individual character. To date, the Japanese government is likely the only one going to such lengths to socially accommodate and integrate AI and robots as a key part of its society's foundation.

(p. 619) Academic Thought and Local Practices: Partnership-Inspired

Takehiro Ohya, professor at Keio University, comments that the view of AI and robots as tools in the Social Principles document is intentional. Ohya explains that the Japanese seem to differentiate between human beings and AI or robots to a lesser degree, which might make them less human-centric, at least in comparison with Western cultures. As a Council member, Ohya shares that he, among other members, view the human-centrist nature of the social principles as a way to better communicate and achieve consensus with Western countries. He also says that he encouraged the Council to consider acknowledging AI and robots as legal persons, since the consideration of them as another species was not deemed ethically justifiable.[48]

The Ethical Guidelines of the Japanese Society for Artificial Intelligence (JSAI) seem to diverge on this point.[49] Article number 9 of JSAI's Ethical Guidelines notes that "AI must abide by the policies described above in the same manner as the members of the JSAI in order to become a member or quasi-member of society."[50] A JSAI blog post explains that the unique article reflects the views of the JSAI and also follows the spirit of Asimov's Three Laws of Robotics. The JASI ethics committee entertained the various ways in which AI and robots would be used by future societies and believes that this communicates the ambitions of Japanese AI and robotics researchers and developers.[51]

These Japanese views and practices build upon the country's long robotic heritage. Yasuo Kuniyoshi, director of the Next Generation AI Research Center at the University of Tokyo, maintains that the Japanese are known for their admiration of hardware, at times over software. As such, he believes that the Japanese are more inclined to trust and appreciated an embodied AI, often in the form of robots, over a bodiless system.[52] Arisa Ema, assistant professor at the University of Tokyo, comments that the Japanese have, at times,

come to view robots as partners due to a long history of robot-friendly popular culture. She also notes, however, that a significant portion of current fascination is generated by the perceived ability of intelligent robots to solve pressing Japanese problems such as care for its super-aging society and automation to revitalize its slowing economy.[53] Another widespread use of AI and robots as tools in Japan is to aid with rescue missions and provide operational assistance in the aftermath of natural disasters.[54]

(p. 620) Popular Culture: Partnership-Rich

Japanese popular culture serves as a key source of inspiration and influence in shaping academic thought and local practices. Numerous Japanese interviewees pointed to two famous cartoons as the source of inspiration for AI and robotics developers. The first is Astro Boy or Mighty Atom (鉄腕アトム), created by Osamu Tezuka and first published in 1952. Atom, a humanoid intelligent robot, was created to replace the son of the Science Ministry's Head, but was discarded when it failed to grow older as a human would. Atom is then sold to the circus but rescued by the new Head of the Science Ministry, who gives robots human rights and builds Atom a humanoid robot family. Atom then goes on to attend elementary school and save the world with its superhuman strength.[55] The story continues to be readapted and entertain audiences today. Jun Murai, co-director of the Cyber Civilization Research Center at Keio University, maintains that many Japanese researchers were influenced by Astro Boy, who introduced them to Asimov's Three Laws of Robotics and the importance and necessity of robot ethics.[56]

The second is Doraemon (ドラえもん), created by Hiroshi Fujimoto and Motoo Abiko (under the pen name Fujiko F. Fujio) and first published in 1969. Doraemon is an intelligent cat robot sent back in time to a Japanese kid named Nobita by his decedent, in the hopes of changing Nobita's lazy behavior. Doraemon and Nobita become close friends and go on adventures across time and space. Doraemon was adapted into numerous animated cartoons and movies and became an internationally beloved character.[57] It has become so popular that Japan's Foreign Ministry named it Japan's first "anime ambassador" in 2008.[58] Hirotaka Osawa, assistant professor at the University of Tsukuba, notes that Doraemon serves as a continued influential icon for AI and robot developers as the character remains relevant and entertaining today still.[59]

Existing somewhere between a tool and a potential partner is Softbank's robot Pepper. Pepper is a conversational humanoid robot with emotional and facial recognition capabilities. Pepper is co-developed with institutions to function as an assistant[60] and even a Buddhist priest.[61] In a Softbank commercial titled "Future Life with Pepper" the company reveals a futuristic vision for the robot acting as a friend, sibling, potential love (p. 621) interest, entertainer, and caretaker.[62] This further blurs the intended use line between a tool and a partner as Pepper functions as an assistant, and priest, but is also explicitly envisioned to become much more than that.

A more partner-oriented practical example is Aibo, the pet robot dog from Sony. Aibo was first available between 1999 and 2006[63] and was officially relaunched in January 2019,

reinforced with machine learning.[64] Jiro Kokuryo, a professor at Keio University, explains that while Aibo robots mimic real dogs, they also have distinct robotic features to avoid uncanny valley-associated fears.[65] In spite of these design measures, when an Aibo was beyond repair, owners sent it off with an elaborate Buddhist burial ceremony, demonstrating a unique attachment to over eight hundred buried robot dogs.[66] Sony, on its part, approaches AI as a tool to achieve "harmony with society" in its Sony Group AI Ethics Guidelines.[67]

The treatment of robots with religious care derives from the concept of animism. Animism can be found in the two major religions in Japan: Shinto and Buddhism. Shinto is a complex, exclusively Japanese religion where the borders between the worldly and otherworldly are blurred. As a polytheistic religion, Shinto belief holds that the spirits of otherworldly beings (e.g., gods) can dwell in animate and inanimate objects, like technology. This suggests a deep spiritual connection between the worldly and manifestation of the otherworldly, often referred to as techno-animism.[68]

In Japanese Buddhism, like the previously discussed Chinese version, both animate and inanimate objects are a part of the natural world and possess the character of Buddha and potential of becoming Buddha. This follows the saying 山川草木国土悉皆成仏, which roughly translates into "all things have the nature of Buddha." This idiom is believed to share the same ancient Indian origins with the aforementioned Chinese one. And in another similar line, a Japanese Buddhist temple also has its own AI robot monk called Android Kannon, which is capable of delivering full Buddhist sermons.[69] As the notion of Japanese techno-animism developed it drew inspiration from both religions, creating a rich synthesis where the source of animation may be different, but the animation of technological artifacts is, in principle, the same.

(p. 622) Much like China and South Korea, Japanese popular culture tells numerous stories about AI and robots as partners and, in particular, love interests. Among them are 絶対彼氏 or Absolute Boyfriend (2008), which was readapted in multiple Asian countries; 僕の彼女はサイボーグ or Cyborg She (2008); キュート or Q10 (2010); イヴの時間 or Time of Eve (2010); ちょびっツ or Chobits (2011); 安堂ロイド～ A.I. knows LOVE? or Ando Lloyd—A.I. Knows Love? (2013). Unlike South Korea and China, however, Japan's mainstream romantic fascination with AI and robots isn't limited to dramas. Vinclu Inc.'s Gatebox AI lab creates a holographic virtual wife and home assistant hybrid modeled after a young female character named Hikari Azuma.[70] The popularity of a virtual wife underscores the pervasive loneliness experienced in Japan,[71] and tells us of what trying to solve social problems with technology already looks like. It also constitutes a rare edge case of intentional tool anthropomorphizing and female objectification, where a functional home assistant is specifically designed to act as a meaningful romantic partner.

Conclusions

Japan is a rich source of information for discussions on perspectives on and approaches to AI and robots as partners. While its policy and social principles may be moving further in-

to the tool range to create international consensus, it is hard to ignore the overwhelming positive approach demonstrated by local culture and practices toward human-AI-robots partnerships. This creates an interesting dual tension between political ambitions and social values. On the one hand, partner AI and robots can prove to be useful tools for a super-aging population, particularly as Japan sees automation as an economic boon, not a threat. On the other hand, this tension does pose other, more complicated "antisocial" questions regarding how this already affects objectified population groups like females, and might affect attempts to repopulate a super-aging country that, simply put, loves AI and robots.

Chapter Conclusions and Discussion

To conclude, South Korea, China, and Japan share considerable similarities despite being placed in three different ranges on the tool-partner spectrum. Each country, in its (p. 623) own way, debates its movement across the spectrum. To date, South Korean policy makes a stand against partner AI and robots while popular culture explores the idea. Chinese policy is headed in the direction of a tool-oriented AI and robotics ethical guidelines, while local practices and culture experiment with the idea of physical and spiritual partnership. Japan's social principles are also moving in the tool direction, but its society actively seeks and creates partner-like AI and robots. As the technology and its widespread societal use continue to develop, we can expect further movement on the spectrum. This movement will surly highlight and plausibly aggravate tensions between top-down tool and bottom-up partner perspectives on and approaches to AI and robots. Sooner or later, these tensions will be at the core of debating the social benefit and harm of AI and robots use.

Three cross-cutting AI and robotics-related ethical issues highlighted by this chapter are female objectification, the Anthropomorphized Tools Paradox, and "antisocial" development. The globally shared issue of female objectification is particularly salient in AI and robotics. Alongside the decisive disenfranchising effect it has on women, it further reinforces the Anthropomorphized Tools Paradox, where functional tools are given desirable, and often female, companionship characteristics to make them more enticing to use. Put together, these two ethical issues create a vicious cycle that subjects both women and technology to the biased objectification of mostly male AI and robots developers and designers. Such development and design choices blur the lines in ways that lead to problematic treatment toward women and even minors, and emotionally and psychologically confuse users. Technology is genderless and artificial. All relevant stakeholders would do well to remember that.

This further underscores the question of "antisocial" technology. What degree of AI and robots' socialization capability development is considered "antisocial"? How many human functions can and should we substitute before we hit that threshold? Could and should the Anthropomorphized Tools Paradox serve as a potential threshold? Despite it being clear that both female objectification and the Anthropomorphized Tools Paradox fall un-

der “antisocial” technology development, they remain incredibly common. We need prosocial regulation of “antisocial” technology if we are seeking to create AI-ready societies. There are many paths to developing and designing AI and robots in ways that do not replace or degrade humans. Human-AI-robot harmony cannot be achieved by creating artificial substitutions to compensate for the fact that we have yet to achieve all-human harmony. There are no technical solutions to social problems.

Bibliography

China

Beijing Academy of Artificial Intelligence. *Beijing AI Principles*, May 28, 2019, **https://baip.baai.ac.cn/en?fbclid=IwAR2HtIRKJxxy9Q1Y953H-2pMHl_bIr8pcsIxho93BtZY-FPH39vV9v9B2eY**.

(p. 624) China Institute for Science and Technology Policy at Tsinghua University. “Chapter 5: Public Perception and General Impact of AI,” in *China AI Development Report 2018*, July 15, 2018. **http://www.sppm.tsinghua.edu.cn/eWebEditor/UploadFile/China_AI_development_report_2018.pdf**.

Japan

Ema, Arisa. *IEEE EAD Regional Reports on AI Ethics: Japan*, April 9, 2018. **http://bai-japan.org/en/2018/reports-on-ai-ethics-japan/**(reprinted from IEEE EAD v. 2).

Robertson, Jennifer. *Robo Sapiens Japanicus: Robots, Gender, Family, and the Japanese Nation*. Oakland: University of California Press, 2018.

South Korea

Statues of the Republic of Korea. *Intelligent Robots Development and Distribution Promotion Act*. **http://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=39153&type=lawname&key=robot**.

Yoo, Juyoung. “Results and Outlooks of Robot Education in Republic of Korea.” *Procedia—Social and Behavioral Sciences* 176 (2015): 251–254.

Additional Resources Reflecting AI Ethics from East Asian Perspectives

Ethically Aligned Design 1st Edition. *Classical Ethics in A/IS*, March 25, 2019. **https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead1e_classical_ethics.pdf**.

Zeng, Yi. *Linking Artificial Intelligence Principles*, 2019. **http://www.linking-ai-principles.org**.

Notes:

([1]) Merriam-Webster Dictionary, *Anthropomorphize*, n.d., https://www.merriam-webster.com/dictionary/anthropomorphize.

([2]) The 2007 and 2016 Robots Ethics Charter iterations published by the Korea Institute for Robot Industry Advancement were kindly translated by Sunyong Byun, professor at the Seoul National University of Education.

([3]) National Information Society Agency (NIA), *Ethics Guideline for the Intelligent Information Society*, April 2018, http://eng.nia.or.kr/common/board/Download.do?bcIdx=20239&cbIdx=62611&fileNo=1.

([4]) *Id*. at 3.

([5]) NIA, *Ethics Charter for the Intelligent Information Society*, June 2018, http://eng.nia.or.kr/common/board/Download.do?bcIdx=20239&cbIdx=62611&fileNo=2.

([6]) Sunyoung Byun, remote interview by Danit Gal, March 7, 2019. For more information, see Sunyoung Byun, Hyunwoo Shin, Jinkyu, Jeong, and Hyeongjoo Kim, "A Study on Necessity of the Charter of Robot Ethics and Its Contents," *Journal of Ethics* 112 (2017): 295–319.

([7]) Tara Francis Chan, "South Korea Will Have 85 Robot Volunteers at the Winter Olympics, Including a Robot Torch Bearer," *Business Insider* (December 6, 2017), https://www.businessinsider.com/south-korea-robots-2018-winter-olympics-2017–12.

([8]) "AI Researchers End Ban after S. Korean University Says No to 'Killer Robots,' " *Reuters* (April 9, 2018), https://www.reuters.com/article/tech-korea-boycott/ai-researchers-end-ban-after-s-korean-university-says-no-to-killer-robots-idUSL8N1RM2HN.

([9]) Ji-hye Jun, "KAIST Launches Ethics Subcommittee on AI," *Korea Times* (June 8, 2018), https://www.koreatimes.co.kr/www/tech/2018/06/129_250278.html.

([10]) Chi Hyung Jeon, remote interview by Danit Gal, March 7, 2019.

([11]) KAIST Institute for Artificial Intelligence, *KAIST Code of Ethics for Artificial Intelligence*, April 4, 2018, https://kis.kaist.ac.kr/index.php?mid=KIAI_O.

([12]) Jaewon Kim, "A 'Lost Generation' in South Korea Bears the Brunt of Rising Inequality," *Nikkei Asian Review* (December 14, 2017), https://asia.nikkei.com/Economy/A-lost-generation-in-South-Korea-bears-the-brunt-of-rising-inequality.

([13]) Jiwon Kim, remote interview by Danit Gal, March 8, 2019.

([14]) Chan Mi Park, Yuin Jeong, Kwangmin Jeong, Hae-Sung Lee, Jeehang Lee, and Jinwoo Kim, "Intelligent Social Robots in the Wild: A Qualitative Study on Deploying Intelligent

Social Robots with Growing Features to Real Home Environments" (February 2018), http://socialrobotsinthewild.org/wp-content/uploads/2018/02/HRI-SRW_2018_paper_4.pdf.

([15]) Hyewon Lee, Hyemee Kang, Min-Gyu Kim, Jaeryoung Lee, and Sonya S. Kwak, "Pepper or Roomba? Effective Robot Design Type Based on Cultural Analysis between Korean and Japanese Users," *International Journal of Software Engineering and Its Applications* 10 (2016): 37–46.

([16]) NIA, *Ethics Guideline*, 3.

([17]) Graham Webster, Rogier Creemers, Paul Triolo, and Elsa Kania, "Full Translation: China's 'New Generation Artificial Intelligence Development Plan," *New America* (August 1, 2017), https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ (last modified October 2018).

([18]) For more information on CAAI, see http://www.caai.cn.

([19]) Phoebe Zhang, "China's Top AI Scientist Drives Development of Ethical Guideline," *South China Morning Post* (January 10, 2019), https://www.scmp.com/news/china/science/article/2181573/chinas-top-ai-scientist-drives-development-ethical-guidelines.

([20]) Celine Ge, "Meet Jiajia, the Realistic 'Robot Goddess' Built by Chinese Researchers," *South China Morning Post* (April 18, 2016), https://www.scmp.com/news/china/society/article/1936834/meet-jiajia-realistic-robot-goddess-built-chinese-researchers.

([21]) Na Chen, "AI Association to Draft Ethics Guidelines," *Chinese Academy of Science* (January 10, 2019), http://english.cas.cn/newsroom/news/201901/t20190110_203885.shtml.

([22]) Masha Borak, "China Wants to Make Its Own Rules for AI Ethics," *ABACUS* (March 4, 2019), https://www.abacusnews.com/future-tech/china-wants-make-its-own-rules-ai-ethics/article/3001025.

([23]) 新浪财经综合, "李彦宏三大提案:完善电子病历 加强人工智能伦理研究," 新浪财经 (March 4, 2019), http://finance.sina.com.cn/review/jcgc/2019-03-04/doc-ihsxncvf9583301.shtml.

([24]) 博客园, "Li Yanhong Unveiled after 'Baidu Lost the Land': 'Simple Search' without Advertisement, Mass Production of Unmanned Vehicles in July," *China IT News 3.0* (May 26, 2018), http://www.fonow.com/view/208592.html.

([25]) 腾讯研究院, "2019年两会，马化腾提了这7份建议案|2万字全文版," (March 6, 2019), https://mp.weixin.qq.com/s/yb6OyuIvHQjImcNMJsJYSw.

([26]) 张志东, "腾讯创始人张志东：信息过载时代，科技如何向善？," 腾讯研究院 (January 20, 2018), https://mp.weixin.qq.com/s/656xTSvP1rLl6SORRS94ew.

([27]) 司晓, “司晓：打造伦理”方舟，让人工智能可知、可控、可用、可靠, 腾讯研究院 (December 6, 2018), https://mp.weixin.qq.com/s/_CbBsrjrTbRkKjUNdmhuqQ [English available].

([28]) ZX, “Beijing Publishes AI Ethical Standards, Calls for Int’l Cooperation,” *Xinhua* (May 26, 2019), http://www.xinhuanet.com/english/2019–05/26/c_138091724.htm?fbclid=IwAR3VPl45lsmmt0anWolOZkpnHpODQa-3bzD9Q8DVB6mZUrB_2xtDxM5vBic.

([29]) Beijing Academy of AI, “Beijing AI Principles,” May 28, 2019, https://baip.baai.ac.cn/en?fbclid=IwAR2HtIRKJxxy9Q1Y953H-2pMHl_bIr8pcsIxho93BtZY-FPH39vV9v9B2eY.

([30]) Miao Liao, remote interview by Danit Gal, March 10, 2019.

([31]) For more information on the 工程伦理 textbook, see: http://www.tup.tsinghua.edu.cn/booksCenter/book_06831902.html.

([32]) Yi Zeng, *Harmonious Artificial Intelligence Principle*, n.d., http://bii.ia.ac.cn/hai/.

([33]) Yi Zeng, remote interview by Danit Gal, March 14, 2019.

([34]) Hanniman Huang, remote interview by Danit Gal, March 12, 2019.

([35]) This animistic inclination is often believed to be a part of other Chinese religions or philosophies, but is more often than not the result of these religions and philosophies melding with the country’s long-standing animistic heritage and tradition.

([36]) Joseph Campbell, “Robot Monk Blends Science and Buddhism at Chinese Temple,” *Reuters* (April 22, 2016), https://www.reuters.com/article/us-china-religion-robot/robot-monk-blends-science-and-buddhism-at-chinese-temple-idUSKCN0XJ05I.

([37]) Jiefei Liu, “Longquan Temple Is Using Artificial Intelligence to Organize and Spread Buddhist Scriptures,” *Technode* (July 9, 2018), https://technode.com/2018/07/09/longquan-temple-techcrunch-hangzhou/.

([38]) Geoff Spencer, “Much More than a Chatbot: China’s Xiaoice Mixes AI with Emotions and Wins over Millions of Fans,” *Microsoft Asia News Center* (November 1, 2018), https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/.

([39]) SNH48 Official Site, 丝芭传媒打造国内首个虚拟偶像组合五月薇 *VIV* 深耕偶像产业 (February 3, 2019), http://www.snh48.com/html/allnews/zixun/2018/0203/2882.html.

([40]) Bernard Marr, “One of the World’s Most Watched TV Shows Will Be Hosted by Artificial Intelligences,” *Forbes* (January 29, 2019), https://www.forbes.com/sites/bernardmarr/2019/01/29/the-worlds-most-watched-tv-show-will-be-hosted-by-artificial-intelligences/#5cd0ce8e68de.

([41]) Cate Cadell, “And Now for Something Completely Different: Chinese Robot News Readers,” *Reuters* (November 9 2018), https://www.reuters.com/article/us-china-tech-ai-

anchor/and-now-for-something-completely-different-chinese-robot-news-readers-idUSKCN1NE19O.

([42]) Government of Japan, *The 5th Science and Technology Basic Plan*, January 22, 2016, https://www8.cao.go.jp/cstp/english/basic/5thbasicplan.pdf.

([43]) *Id*. at 13.

([44]) Cabinet Office Council on the Social Principles of Human-centric AI, *Social Principles of Human-centric AI (Draft)*, n.d., http://search.e-gov.go.jp/servlet/PcmFileDownload?seqNo=0000182653.

([45]) *Id*. at 1.

([46]) *Id*. at 4.

([47]) *Id*. at 6.

([48]) Takehiro Ohya, remote interview by Danit Gal, March 7, 2019.

([49]) Japanese Society for Artificial Intelligence (JSAI), *The Japanese Society for Artificial Intelligence Ethical Guidelines*, May 16, 2017, http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf.

([50]) *Id*. at 3.

([51]) JSAI, *About the JSAI Ethical Guideline*, February 28, 2017, http://ai-elsi.org/archives/514.

([52]) Yasuo Kuniyoshi, remote interview by Danit Gal, March 11, 2019.

([53]) Arisa Ema, remote interview by Danit Gal, March 11, 2019.

([54]) Martin Fackler, "Six Years after Fukushima, Robots Finally Find Reactors' Melted Uranium Fuel," *New York Times* (November 19, 2017), https://www.nytimes.com/2017/11/19/science/japan-fukushima-nuclear-meltdown-fuel.html.

([55]) For more information on Astro Boy, see: http://tezukainenglish.com/wp/?page_id=138.

([56]) Jun Murai, remote interview by Danit Gal, March 8, 2019.

([57]) Yasuyuki Yokoyama, "Celebrating Exactly 100 Years before Doraemon's Birthday," *Nippon.com* (December 10, 2012), https://www.nippon.com/en/currents/d00056/celebrating-exactly-100-years-before-doraemon's-birthday.html.

([58]) "*Doraemon Named 'Anime Ambassador,'" Japan Today* (March 17, 2008), https://japantoday.com/category/entertainment/doraemon-named-anime-ambassador.

([59]) Hirotaka Osawa, remote interview by Danit Gal, March 8, 2019.

([60]) For more information on Pepper, see https://www.softbankrobotics.com/emea/en/pepper.

([61]) Alex Martin, "Pepper the Robot to Don Buddhist Robe for Its New Funeral Services Role," *Japan Times* (August 16, 2017), https://www.japantimes.co.jp/news/2017/08/16/business/pepper-the-robot-to-don-buddhist-robe-for-its-new-funeral-services-role/.

([62]) "Future Life with Pepper," YouTube video, posted by "Dean Khaled," February 29, 2016, https://www.youtube.com/watch?v=-A3ZLLGuvQY.

([63]) Sony-Aibo, *History*, n.d., http://www.sony-aibo.co.uk/history/.

([64]) "Sony Starts Taking Advanced Orders for New Version of Aibo Robot Dog," *Japan Times* (July 19, 2018), https://www.japantimes.co.jp/news/2018/07/19/business/tech/sony-starts-taking-preorders-new-version-aibo-robot-dog/.

([65]) Jiro Kokuryo, remote interview by Danit Gal, March 3, 2019.

([66]) Suzuki Miwa, "In Japan, Aibo Robots Get Their Own Funeral," *Japan Times* (May 1, 2018), https://www.japantimes.co.jp/news/2018/05/01/national/japan-aibo-robots-get-funeral/#.XJt2rvwRV0t.

([67]) "Sony Group AI Ethics Guidelines," Sony (September 25, 2018), https://www.sony.net/SonyInfo/csr_report/humanrights/hkrfmg0000007rtj-att/AI_Engagement_within_Sony_Group.pdf.

([68]) Casper Brunn Jensen and Anders Blok, "Techno-animism in Japan: Shinto Cosmograms, Actor-network Theory, and the Enabling Powers of Non-human Agencies," *Theory, Culture & Society* 30 (2013): 84–113.

([69]) Thisanka Siripala, "An Ancient Japanese Shrine Debuts a Buddhist Robot," *The Diplomat* (March 5, 2019), https://thediplomat.com/2019/03/an-ancient-japanese-shrine-debuts-a-buddhist-robot/.

([70]) For more information, see https://gatebox.ai/home/ and https://gatebox.ai/news/2018/07/31/01/ [in Japanese].

([71]) Michael Hoffman, "Japan Struggles to Keep Loneliness at Arm's Length," *Japan Times* (November 10, 2018), https://www.japantimes.co.jp/news/2018/11/10/national/media-national/japan-struggles-keep-loneliness-arms-length/.

Danit Gal

Danit Gal, Technology Advisor to the UN Secretary General High-Level Panel on Digital Cooperation

MIT Technology Review
Insights / Research

Asia's AI agenda

The ethics of AI

A research program in association with:

Founding sponsor

ADP

Gold sponsor

SG:D
SINGAPORE
EMPOWERING POSSIBILITIES

Silver sponsors

GENESYS

splunk>

ASIA School of Business
in collaboration with MIT Sloan Management

Preface

Asia's AI agenda is an MIT Technology Review Insights research program sponsored by ADP, Infocomm Media Development Authority, Genesys, Splunk, and the Asia School of Business. It is designed to comprehensively examine the development of artificial intelligence (AI) in Asia Pacific from four distinct angles: Asia's AI ecosystem, the leading use cases and business applications across the region, the evolving talent landscape, and the emerging discussions around AI and ethics. To produce this series, MIT Technology Review Insights conducted a survey of almost 900 executives across 13 markets, and a series of interviews with leading authorities from academia and industry. The research is editorially independent and the views expressed are those of MIT Technology Review Insights.

The survey

- In September and October 2018, MIT Technology Review Insights surveyed 871 senior business leaders, of whom 29 % are CIOs, CTOs, or heads of AI or analytics. More than half (53%) are C-suite and director-level executives. Almost 60 % are from large companies with over $1 billion in revenue.
- Survey respondents are based in 13 Asia-Pacific economies: Australia, China, Hong Kong, India, Indonesia, Japan, Malaysia, New Zealand, the Philippines, Singapore, South Korea, Thailand, and Vietnam, with a minimum of 50 responses from each.
- Respondents are drawn from a wide range of industries, including more than 50 from each of the following sectors: consumer goods and retail; financial services; information technology and communications; manufacturing, pharmaceuticals and healthcare; professional services; property, construction, and engineering; and transport and logistics.

Expert interviews

We would like to thank the following experts for contributing their time and insights towards this research program:

Goh Yihan, Dean, School of Law, **Singapore Management University**, Singapore

Loredana Padurean, Associate Dean, **Asia School of Business**, Malaysia

Tomaso Poggio, Eugene McDermott Professor, Department of Brain and Cognitive Sciences, **MIT**, United States

Kenji Suzuki, Professor, Center for Cybernics Research and Faculty of Engineering, Information and Systems, **University of Tsukuba**, Japan

Viriya Upatising, Chief Information Officer, **True Corporation**, Thailand

Toby Walsh, Scientia Professor of Artificial Intelligence, **University of New South Wales**, Australia

Ken Wong, Head of AI Lab, **OCBC Bank**, Singapore

Xue Lan, Professor, School of Public Policy and Management, and Dean, Schwarzman College, **Tsinghua University**, China

Zee Kin Yeong, Assistant Chief Executive, Data Innovation and Protection Group, **Infocomm Media Development Authority**, Singapore

Contents

1. Executive summary 4
2. Introduction: Who's afraid of AI? 5
   - The next space race 5
3. AI battlefields 7
   - Who, me? 7
   - Malicious use and weaponization 9
   - Inequality and unemployment 9
4. Emerging governance for AI 11
   - New institutions and global partnerships 11
   - Codes and principles 12
   - Companies and the AI ethics debate 14
5. Challenges for AI ethics: the view from the private sector 15
   - AI is not yet replacing workers outright 15
   - There is no "global" 15
   - Unintended consequences 16
6. Conclusion 17

1. Executive summary

Globally, future outlooks for artificial intelligence (AI) swing between two extremes—excited anticipation about the positive impact AI will have on economies and societies, and deepening fear of its potential to disrupt livelihoods and do harm. In Asia, governments and civil society groups are concerned about defining regulatory frameworks to guard against the latter, and all ecosystem experts are grappling with how to steer AI toward the former, more socially advantageous directions. On balance, however, the Asian business leaders surveyed for this report have great optimism about AI's positive effect on their businesses, societies, and individual well-being.

This report, combining an Asia-wide executive survey with expert interviews from industry, government, and academia, takes the pulse of public and private actors in the AI ethics debate in the region. The key findings of the report are:

- **AI will be a major growth driver for Asia in the coming decade.** The company priorities for AI are to enhance customer satisfaction, speed up decision-making, and reduce inefficiencies. The loss of some roles to automation, and the restructuring of others to take advantage of technology-created capacity, are likely. Yet reducing headcount is not a top priority in and of itself. Just one-third of survey respondents listed the need to reduce labor costs as a top-three driver for AI.

- **Biases within AI tools are potentially dangerous for Asia—but biases about AI's use in Asia could be even more so.** Asia's AI ecosystem participants are aware of and concerned about the potential for embedded biases (race, gender, or socio-economic status) within AI tools, and the harm this can cause through facilitating over-policing of minority communities, or economic exclusion. Weaponization and malicious use of AI are also ethical concerns in Asia as applications are increasingly commoditized and industrialized. While Asian decision-makers are concerned about a potentially negative impact, particularly where jobs are concerned, optimism is the more dominant sentiment, which will propel the use of AI in Asia.

- **Asian governments are building institutional capacity and frameworks to increase AI governance—but have yet to develop regulations.** Overwhelmingly, more survey respondents believe Asia will lead the world in the development of ethics and governance than any other region: 45%, as compared to only a quarter who see North America as the ethics front-runner. Across the region, from Singapore to Japan and China, governments are assembling AI institutions to guide governance, often consulting with the private sector and civil society.

- **Asian respondents are engaged in AI ethics discussions and see a constructive role for governments.** Just under half (42%) of participants say there is "vigorous debate" on ethical issues surrounding AI in their company, and the majority (55%) think AI should be government-regulated. Both Asia's governments and its businesses are committed to maximizing AI's benefits, so the fact that talent availability is the region's top AI deployment challenge (according to 58% of respondents) suggests that those with AI skills will wield significant bargaining power in the debate which lies ahead.

- **AI-driven unemployment narratives are counterbalanced by the potential to enhance and augment human work.** Many survey respondents share one of the world's biggest ethical fears—that the unchecked use of AI will result in massive loss of jobs and livelihoods; 42% believe that the rise of AI in Asia will destroy more jobs than it will create. However, other survey responses reveal that Asian business leaders are not overly concerned about job loss in their own organizations, and they believe that AI will benefit their employees. The majority (59%) of respondents believe their employees' job roles have been enhanced since the introduction of AI.

2. Introduction: Who's afraid of AI?

AI technology is far from new, but its capability has leaped forward over the past half decade, allowing it to move from the margins to the mainstream. AI is increasingly becoming industrialized and commoditized in areas ranging from logistics and health diagnostics to personal assistants and vehicles. Across Asia, the ecosystem is responding to watershed moments in global AI development with competitive zeal and redoubled efforts.

The AI shake-up of Go, the adversarial board game known for its momentous complexity, is one such recent example. When AlphaGo, a program developed by Alphabet's AI company DeepMind Technologies, defeated Korean international Go champion Lee Sedol in March 2016, it was considered a singular achievement. When a successor program, AlphaGo Master, defeated China's world champion Ke Jie a year later in May 2017, it set off what has been described as China's "Sputnik moment" (referring to the panic set off in the U.S. when the Soviet Union launched the first satellite into orbit in 1957, spurring the space race).

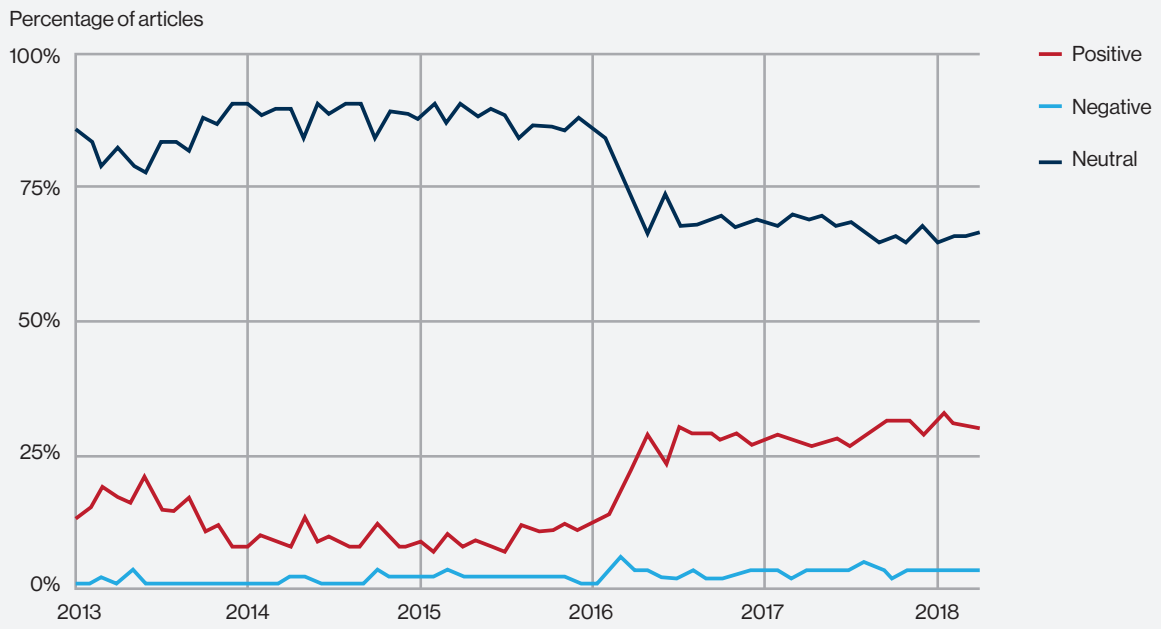
The next space race

China's accelerated AI development plan, known as the "AI 2030 agenda," was released the following month, and articulated an overarching national vision for the numerous technology initiatives already underway. Among these was a homegrown response to AlphaGo: by January 2018, a Go program called Fine Art, developed by Chinese internet giant Tencent, had also bested Ke Jie.

While media coverage of AlphaGo and Fine Art, like most stories on AI over the past few years, reported on the victories of computers over humans, articles written about AI since 2016 have become markedly more positive in their sentiment about the technology.[1]

At the same time, a recent World Economic Forum survey of public and private sector leaders

Figure 1: Sentiment analysis of articles referencing AI, 2013-18



Source: TrendKite, Artificial Intelligence Index 2018[1]

[1] *cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf*

found that AI and robotics were seen as the emerging technologies with the greatest potential negative impact (see Figure 2).[2] This shows that while attitudes towards AI are generally improving, there are still grave concerns about the knock-on effects.

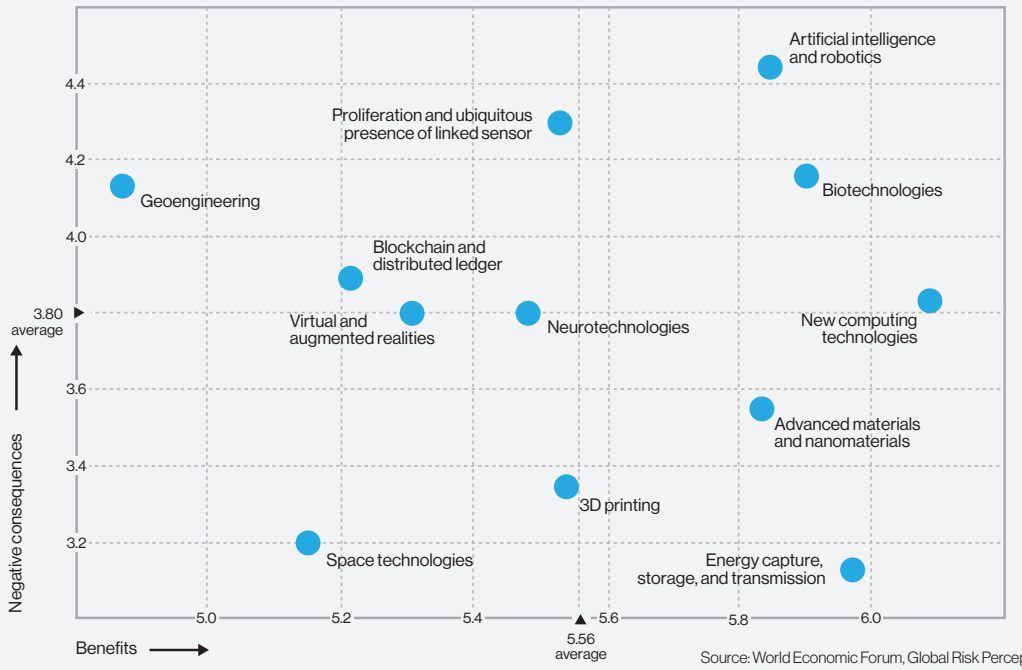
"The first car was invented in the late 1800s, but it took a century for our use of automotive technology to fully mature," says Loredana Padurean, associate dean and faculty director for action learning at the **Asia School of Business**, a collaboration between the Central Bank of Malaysia and MIT's Sloan School of Management. "AI is advancing at 10 times that speed—AI is going to advance faster than any other innovation we have ever seen," she says, noting the insufficient time for debating the potential impact on our societies and economies.

The risks of AI and autonomous systems are not hypothetical. Biases, accidents, and controversial uses (such as AI-enabled weapons and other military applications) have all been features of the AI landscape over the past two years. Asian government and business leaders are not blind to these risks. However, as this report will argue, AI ecosystem participants are more sanguine about the "upsides" of AI, and are confident that Asia has the will, the resources, and the institutional capacity to develop robust ethical frameworks. Asia's potent cocktail of pro-AI enthusiasm and pragmatism may create further ethical conundrums, but this report argues the region has experience in using technology to increase social harmony and economic sustainability.

"It took a century for our use of automotive technology to mature. AI is advancing at 10 times that speed—AI is going to advance faster than any innovation we have ever seen."

Loredana Padurean
Associate Dean
Asia School of Business

Figure 2: Perceived benefits and negative consequences of 12 emerging technologies



Source: World Economic Forum, Global Risk Perception Survey, 2017

[2] *reports.weforum.org/global-risks-2017/part-3-emerging-technologies/3-1-understanding-the-risk-landscape/*

3. AI battlefields

Bias is a prominent worry in the development of AI because algorithms discriminate in order to make decisions, and those bases are literally "hard-coded" by technical experts. The fact that men account for a much larger share of the AI workforce than women creates the potential for gender biases in algorithm development.

The World Economic Forum "Global Gender Gap Report, 2018" warned that, with women accounting for 22% of AI roles globally, machine learning technologies are being encoded with perspectives that are intrinsically non-representative of the societies they serve.[3] The report observed that in the three top-ranked economies by AI talent, the United States and India had significant gender skews towards male programmers (70 and 72% respectively), and Germany, at 82%, had one on the world's widest gaps. At 61%, Singapore, the report's eighth-ranked AI economy, was the only Asian country in the top 20 to have a substantially smaller gender gap.

In the July 2018 "China AI Development Report", the China Institute for Science and Technology Policy at Tsinghua University analyzed the results of an "AI interest" survey of users of Toutioa, a domestic news-aggregation platform, finding that throughout 2017, 66% of all views of AI-related articles were by men.[4] Despite efforts to bring women into scientific and technical careers, it is clear that more still needs to be done.

Figure 3: Share of professionals with AI skills, by gender and country

Country rank, AI skills	Country	Female (%)	Male (%)
1	United States	23	77
2	India	22	78
3	Germany	16	84
4	Switzerland	19	81
5	Canada	24	76
6	France	21	79
7	Spain	19	81
8	Singapore	28	72
9	Sweden	20	80
10	United Kingdom	20	80
11	Netherlands	21	79
12	Poland	16	84
13	Australia	24	76
14	Brazil	14	86
15	Italy	28	72
16	Turkey	24	76
17	Belgium	19	81
18	South Africa	28	72
19	Mexico	15	85
20	Argentina	17	83

Source: World Economic Forum, Global Gender Gap Report, 2018

Who, me?

Race is another bias that can AI can learn from its developers. In the United States and Europe, facial recognition has proven less effective in identifying non-white faces, and researchers have shown that some such tools are more likely to misidentify minorities as criminals. Such systems are already in use by many police forces globally, giving rise to fears that racial biases in AI-enabled facial recognition can lead to "false positives" and wrongful arrests. California's high-tech heartlands of San Francisco and Oakland may soon pass legislation banning the use of facial recognition

[3] *www3.weforum.org/docs/WEF_GGGR_2018.pdf*
[4] *www.sppm.tsinghua.edu.cn/eWebEditor/UploadFile/China_AI_development_report_2018.pdf*

and other digital surveillance tools by their municipal governments.[5] The need to guard against bias intensifies as companies move into sensitive sectors, such as justice and the law. "You have more legal tech startups coming into the picture using AI, but some of these startups are not even fronted by legally trained people," warns Goh Yihan, dean of the **Singapore Management University** School of Law.

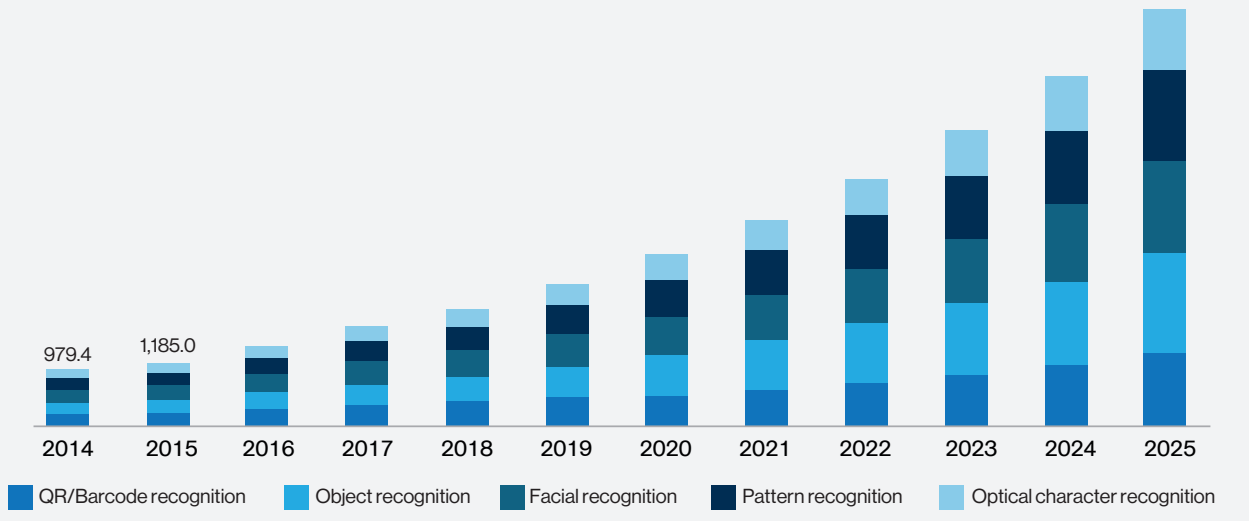
Nevertheless, Asia is likely to become a global frontrunner in the use of facial recognition technology in the coming decade. Market research firm Statistics Market Research Consulting predicts that global facial recognition technology sales will more than triple over the coming decade, from $4 billion in 2017 to nearly $15 billion in 2026.[6] Analysts estimate that half of global demand comes from Asia, and China specifically is recognized as both the world's single largest market and the leading producer of facial recognition technologies.

Since 2015 China been building the world's largest face recognition database, which by 2020 is intended to contain the biometric details of all its 1.3 billion citizens, and a growing body of evidence suggests that biases are being deployed with intent. A recent New York Times report claimed Chinese authorities are using facial recognition to identify and track ethnic Uighurs in cities outside of their traditional home province of Xinjiang.[7] A Dutch cybersecurity researcher, while identifying security breaches in Chinese facial image databases, recently discovered a cache of personal data and images on 1.8 million Chinese women being used to determine their so-called 'BreedReady' status.[8]

Other efforts to control Chinese citizens on socio-political, rather than criminal, grounds include its social credit system that uses AI, big data, and facial recognition to profile and place restrictions on those who fail to pay debts or flout other rules. China's world-leading image-recognition market raises profound questions about not only the intended use, but also the manner in which the technology is being created.

While few Asian governments have China's resources or careful orchestration between the state and industry, most do share a heightened concern for public security and social harmony. AI can, of course, foster public safety and improve resource allocation, but China's example reveals how quickly its application can stray into reinforcing political, ethnic, or ideological biases in the name of ethical goals.

Figure 4: China image recognition market size, by technique, 2014-25 (US$ m)



Source: Grand View Research, 2018

[5] *www.slate.com/technology/2019/05/san-francisco-first-city-ban-facial-recognition.html*
[6] *www.reuters.com/brandfeatures/venture-capital/article?id=48514*
[7] *www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html*
[8] *www.theguardian.com/world/2019/mar/11/china-database-lists-breedready-status-of-18-million-women?CMP=twt_gu*

Malicious use and weaponization

Dystopian scenarios involving AI typically revolve around what would occur should powerful tools fall into the wrong hands. A machine learning system recently developed by academic researchers at the University College of London and the Alan Turning Institute re-identified anonymized Twitter users through meta-data analysis with 96.7% accuracy.[9] If hackers can develop or appropriate such models themselves, the implications for personal data privacy are severe.

Even more severe are the implications if bad actors take over AI-controlled weapons. Lethal autonomous weapons (LAWs) are systems which can, of their own volition, identify a target and complete a lethal attack without human approval or authorization. There is no evidence that a LAW yet exists, but there are technological precursors in several Asian countries, including South Korea and China. While there is currently no international consensus on how LAWs fit into frameworks governing war crimes and accountability, Human Rights Watch believes LAWs contravene the Geneva Convention Martens Clause, which requires emerging technologies not covered by existing treaties to be judged by the "principles of humanity" and the "dictates of public conscience".[10]

Dangerous technologies, from guns to nuclear reactors, can be controlled in a way that AI cannot. "All sorts of algorithms could be repositioned by a person with a bad intent" says Toby Walsh, scientia professor of computer science and engineering at the **University of South Wales**, Sydney. He points to possibilities for radar-linked AI that identifies pedestrians in autonomous-vehicle technology being programmed for drones to hunt targets. Mitigating the risk of malicious AI use is not, he says, possible through out-and-out bans: "You're not going to stop AI technology, just like we didn't stop chemistry by banning chemical weapons, and we didn't stop biological research by banning biological weapons," he says. That said, there is evidence that the AI ecosystem has other means to limit engagement in weaponized AI and lethal weapons. Walsh, along with a coalition of over 50 AI academics, used a boycott to get the Korea Advanced Institute of Science and Technology to agree not to develop AI weapons as part of a collaboration with a defense company.

> "You're not going to stop AI technology, just like we didn't stop chemistry by banning chemical weapons, and we didn't stop biological research by banning biological weapons."
>
> **Toby Walsh**
> **Scientia Professor of Computer Science and Engineering**
> **The University of South Wales**

Malicious AI is also playing out in social media, through disinformation and bot-driven "drown out" campaigns. These reportedly played out in Saudi Arabia in the aftermath of the disappearance of journalist Jamal Khashoggi, and Myanmar's government has been accused of leveraging disinformation to kindle religious and ethnic conflict and antipathy against the Rohingya. Non-state groups also stand accused. In Indonesia, police claim to have uncovered a clandestine fake news operation geared to destabilize its government.

Inequality and unemployment

As bots carry out more complex knowledge tasks in both white- and blue-collar work, there is concern that AI will split labor pools into a "winning" class of workers—whose roles and responsibilities are augmented by AI—and a losing class whose jobs are replaced by automated tools or eliminated entirely. Overall, AI will affect one in every five Asian jobs, and automation will eliminate one in eight, according to the MIT Technology Review Insights report "AI and human capital".[11] The report also found that emerging Asian economies, which are built on labor-intensive industries and services

[9] *www.wired.co.uk/article/twitter-metadata-user-privacy*
[10] *www. treaties.un.org/doc/Treaties/1983/12/19831202%2001-19%20AM/XXVI-2-revised.pdf*
[11] *www.insights.techreview.com/asias-ai-agenda-ai-and-human-capital/*

from electronics and textiles to business process outsourcing, are much more at risk of job losses due to AI and automation than their wealthier counterparts—even though greater percentages of "automatable" jobs will disappear from the workforces of rich Asian countries. Poorer Asian economies are proportionately more dependent on the lower-skilled job classes endangered by AI and, unlike their wealthier peers, have fewer high-skill jobs that will be enhanced by AI, and less institutional capacity to retrain and reskill workers.

AI's potential threat to jobs is not lost on our survey respondents. Some 42% think AI will either destroy or disintermediate more jobs than it creates in Asia and only 19% disagreed. These concerns are filtering down to front-line staff in many Asian organizations. "'Is my job going away because of AI?' is a common question we are asked by staff in nearly every town hall meeting," says Virya Upatising, chief information officer at **True Corporation**, the Thailand-based telecommunications firm.

There is also growing anxiety in China, despite the country's overall culture of techno-optimism, according to Xue Lan, dean of Schwarzman College at **Tsinghua University**, who led the team responsible for the China AI Development Report. "There have been many heated debates in China about the potential impact of AI on employment. In our report we found that, in general, the public is very supportive of AI development, but they are simultaneously concerned about AI's impact on their jobs". Xue notes that the potential impact of AI is a subject to discussion in the general public and also by professional societies, leading to the creation of research groups and organizations.

The majority of Asia-based respondents to MIT Technology Review's survey agree with positive statements about the impact of AI. Four-fifths believe AI will increase productivity overall, 59% believe their employees have had their roles enhanced by AI, and nearly half believe that overall employee satisfaction is higher since the introduction of AI. It is possible that these positive views are borne out of necessity or an "anchoring bias"; some 70% of respondents say that AI is essential to their future competitiveness.

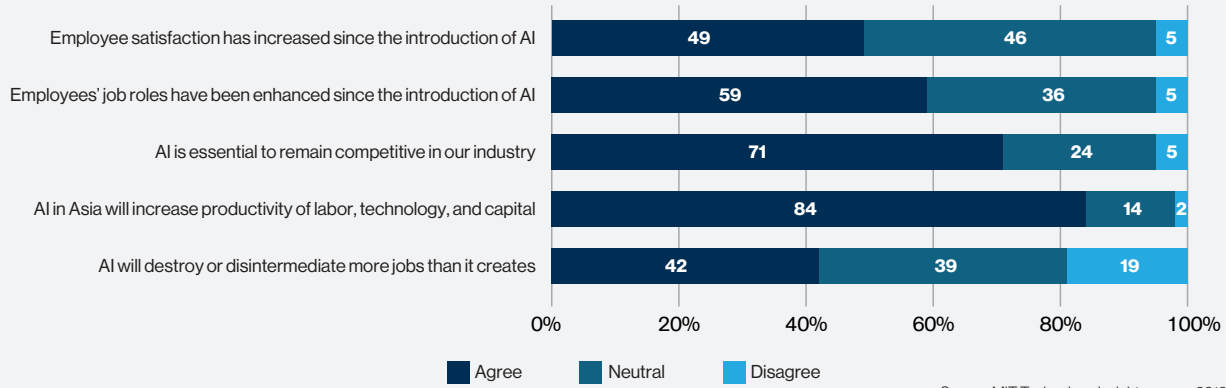
Key takeaway

Race and gender bias—Improving the race and gender diversity of Asia's AI developer pool is critical for mitigating biases - however genuinely unintended.

The all-seeing eye—China's facial recognition lead will be of benefit to some but not others. There will be countless new use cases boosting efficiency, security, and public safety. The price will be diminished civil liberties.

Jobs will go—Only 19% of survey respondents in Asia do not believe that AI will destroy or disintermediate more jobs than it creates.

Figure 5: Perceived impact of AI in Asia (% of respondents)



Source: MIT Technology Insights survey, 2018

4. Emerging governance for AI

To date, the AI-ethics debate has been largely diagnostic: identifying the risks to individuals and societies posed by current or future technologies. Concrete responses from governments and AI technology companies are at a nascent stage. Yet with Asia likely to be at the forefront of AI development and adoption globally over the coming decade, it is also likely that Asia will run into ethical issues and dilemmas early on. Nearly half of survey respondents believe Asia will lead the development of ethics and governance frameworks, compared to only 25% seeing North America as an ethics leader. The real challenge for Asia's economies will be to temper the tendency to rush the deployment of AI to overcome critical infrastructure gaps and to increase citizen welfare with a careful examination of the potential risks to society that may be created as a result.

Concrete responses to today's AI ethics debate from governments and companies are at a nascent stage. Yet with Asia likely being at the forefront of AI development globally over the coming decade, it is also probable that the region will run into ethical dilemmas early on.

New institutions and global partnerships

Early movers in convening expertise and coordinating policy positions are to be found in Europe. The UK government's Centre for Data Ethics and Innovation launched in 2017 and Germany followed with the Data Ethics Commission the following year. Canada and France have formed an international coalition on AI ethics, set to include other Group of Seven and EU members. But Asian governments are not far behind in these efforts. Singapore formed the Advisory Council on the Ethical use of AI and Data in 2018, with advisory members drawn from industry, government, and civil society. The Indian government has tasked its agencies, notably NITI Aayog, to take up the policy work program for AI and publish guidance. Chinese authorities convened an ethics committee in 2018, led by leading AI scientist Chen Xiaoping, and senior figures have called for more international institutional collaboration.

Asian universities are also working closely with governments and global partners on ethics and policy. The Hong Kong University of Science and Technology (HKUST) was the first Asian university to join the "Partnership on AI to Benefit People and Society", an international alliance of more than 80 companies, that has a number of goals for shaping the development of AI in aspirational and benevolent ways. HKUST's addition has significance, given that Sino-US technology collaboration has slowed in the mounting trade war, and is a reminder of Hong Kong's unique position as a transition node for transferring technology and capital between the two superpowers. There are now many university-led AI ethics initiatives, including New Zealand's University of Otago forming a Centre for Artificial Intelligence and Public Policy to work with the ministry of broadcasting, communications and digital media, and the government's digital services division on policy options.

Codes and principles

Goals and broad ethical principles are beginning to be shaped. Policy documents include Japan's AI Technology Strategy (March 2017), China's Next Generation Plan (July 2017), and India's #AIforAll national strategy (June 2018), with key principles including ensuring AI has broad benefits for the nation's development. The Australian government recently earmarked AU$ 29.9m (US$ 20.6m) for a four-year AI and machine learning program, including the development of an "AI Ethics Roadmap". Malaysia's government has announced the development of a national framework for AI to be completed by the end of this year. Most of these frameworks only address the ethical issues in broad terms, and largely informed by pragmatic concerns around job loss and reskilling requirements. The Malaysia Digital Economy Corporation (MDEC), the country's technology promotion and investment coordination body, which is leading the AI framework development, is using it in large part to coordinate knowledge-sharing and best-practice development between Malaysia's AI ecosystem participants (academia, government bodies, and established and startup enterprises) to build coordinated responses to the implications that AI will have on skills, livelihoods, and economic competitiveness.

Other ethical framework implementation efforts are more incentives-driven: Australia's chief scientist Alan Finkel recently proposed that his

Making ethics frameworks actionable

Experts believe the plethora of government guidelines need to balance detail and specificity with flexibility to evolve as technologies do. "Rather than add to existing pool of ethical principles that have been generated by governments, we thought it would be more useful to produce a framework to translate ethical principles into things companies can do in the office," says Zee Kin Yeong, assistant chief executive of the **Infocomm Media Development Authority** (IMDA) of Singapore, and deputy commissioner at the Personal Data Protection Commission. "How do we ensure that ethical principles are properly incorporated into the company's internal governance framework as part of their risk assessments, as part of their code of ethics? And how do we make use of existing corporate governance structures to ensure that there are proper decisions being made at the right level, proper monitoring to identify any incidents, and proper ways of figuring out what are the best ways of using AI to help make decisions within the company?"

IMDA has released a model framework for public consultation to help organizations ensure that decisions made by, or with assistance from, AI are explainable, transparent, and fair to consumers, and that AI solutions are "human-centric". The model framework, in many ways, mirrors government oversight used in corporate governance generally. For instance, it suggests that organizations using AI need to allocate roles and articulate responsibilities for assessing the risks involved in AI adoption and managing the AI model adoption process, as well as reviewing outputs and ensuring effective transparency and customer communication.[15]

In its launch, the IMDA argued that such a framework would enhance trust in, and understanding of, AI and drive acceptance of how decisions are made. Building on a June 2018 discussion paper by the country's data protection commission and IMDA, it maps four ethical principles applying to AI deployment in four areas: internal governance, risk management in autonomous decision-making, operations management, and customer relationship management.

Goh Yihan, dean of the Singapore Management University School of Law, emphasizes the need for frameworks to be agile and evolving. "There is no point in having a framework that is so detailed but is also inflexible and doesn't react to the fast-moving pace of technology. A framework needs to be easily comprehensible. If you have academics write the framework with lawyers, it may run into complex language, and that's not helpful because AI and algorithms are a tool used not only by industry or certain professions, but across different industries."

government implement a certification process to award firms that demonstrate they have adopted AI in a responsible manner, in terms of job-loss management and sustainable manufacturing process. Finkel calls it the Turing Certificate—after the computer scientist (and recognized "founding father" of AI) Alan Turing—and believes that making it mandatory for firms seeking government contracts will send a strong signal to both the Australian and global economy.[12]

Despite the region's prodigious road-mapping, however, articulated positions on ethics are lagging in Asia, relative to other regions, and in terms of other AI development priorities. In an analysis of 18 national and regional AI strategic plans, the Canadian Institute For Advanced Research found that the strategies coming from Asia placed relatively lower priority on ethics than those developed in other regions, although in one related area—the use of AI to promote social and economic inclusion—India scored highest globally.[13] The heat map shows that Asian economies are still largely focused on pointing AI at industrial development strategies.

Along with the ethical chapters in national strategies, governments are also issuing codes and charters, although most lack legal status. In 2012, South Korea issued a "Robot Ethics Charter", which covers standards, illegal use, and data protection. These are written as guidelines rather than laws. Japan issued a "Robot Strategy" three years later (2015) covering policy, ethics, and safety standards.

Figure 6: Heatmap of national AI strategy priorities

Source: CIFAR, 2018

[12] *www.governmentnews.com.au/government-should-lead-ai-certification-finkel/*
[13] *www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf*
[14] *www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf*

China's "Next Generation AI Development Plan", published in July 2017, envisions AI "to improve social management capacity" and pledges research on civil and criminal liability, privacy and IP, information safety, accountability, design ethics, risk assessment, and emergency responses, and commits to participate in AI global governance.

In 2018, Singapore's monetary authority introduced principles to promote fairness, ethics, accountability, and transparency (FEAT) in AI and data analytics in finance. The FEAT Principles require firms to demonstrate that their use of AI Data Analytics (AIDA) does not result in systematic disadvantages for individuals; that analytic models and frameworks are regularly reviewed to reduce or remove biases, and that any decisions based on AIDA are held to the same codes of conduct and same level of ethical scrutiny as would be human-driven decisions.[15]

Companies and the AI ethics debate

Ethical AI is not just a question for government regulators; companies are also part of the debate and potentially the solution, through participation in international fora, and their internal development agendas. Some 42% of Asian business leaders in our survey indicated that there was "vigorous debate" around ethical questions in their companies. Evidence to date shows their activities have included entering international networks, advocating for rules and clarity from government, developing internal codes, and embarking on ethical and pro-social AI product development.

Asian firms have become more active participants in international ethics arenas. Baidu became the first mainland Chinese representation in the Partnership on AI, an alliance with stakeholders including Apple, Google and Amazon. Fujitsu Laboratories of Europe, part of Japan-based Fujitsu, is one of the founding partners of the AI4People initiative, a European forum on the social impacts of AI.

Yet survey data shows that Asian companies are looking to government to take the lead—the majority think it should be government-regulated. Chief executives of Baidu and Tencent, Robin Li and Pony Ma, have recently stated that China must develop guidelines in areas that raise ethical questions, such as autonomous vehicles, gene editing, and data privacy.[16] Li advocates government-led research into ethical AI, telling a meeting of the Chinese People's Political Consultative Conference that "only by establishing a sound set of ethical norms … can we reap more benefits from AI."

There are numerous examples of corporate AI guidelines: Microsoft has a research program called FATE (Fairness, Accountability, Transparency and Ethics), exploring how to combine AI innovation and ethical use; Google's DeepMind has formed an Ethics and Society division; Sony has issued a seven-point AI ethics guideline.[17] Asian companies, or global firms with an Asia presence interviewed for this study, indicate they are developing their own. Genesys, one of the sponsors of this report, has a five-point guide to AI ethics covering transparency, fairness, accountability, data protection, and social benefit. This is not merely a public relations exercise. The firm predicts that by 2021, algorithm opacity, decision bias, malicious use of AI, and data regulations will result in a doubling of spending on relevant governance and compliance staff.

Key takeaway

Frameworks over enforcement—Asian governments and businesses are becoming more active in setting goals and guidelines for the ethical development of the AI industry. Yet thus far, there are no oversight or enforcement mechanisms. Rather than tackling ethics head on and providing solid legal backing and recourse, Asian policymakers are fostering trust between consumers, users of AI, and AI developers so that the industry as a whole can grow.

[15] *www.mas.gov.sg/-/media/resource/news_room/press_releases/2018/Annex%20A%20Summary%20of%20the%20FEAT%20Principles.pdf*
[16] *www.abacusnews.com/future-tech/china-wants-make-its-own-rules-ai-ethics/article/3001025*
[17] *www.microsoft.com/en-us/research/group/fate/*

5. Challenges for AI ethics: the view from the private sector

Building robust and comprehensive responses to the ethical challenges, real and hypothetical, raised by AI is far from straightforward.

AI is not yet replacing workers outright

While it is clear that AI is assisting companies in making routine tasks more efficient, the link between job losses and AI is still opaque. Moreover, should we not grasp at opportunities to pass off low-value work so that humans can focus on higher-order cognitive tasks? One executive interviewed for this report asked "How ethical is it to keep people doing menial tasks when they could actually be taking a step up, when computers can do this for them?"

It depends on how we think about the workforce as a whole, argues Loredana Padurean at the Asia School of Business, rather than a simple question of whether certain jobs will come or go. "If we do it well, AI can improve our life dramatically, because we remove people from working in toxic or demeaning environments. The question is, what do we do with these people? Can we create societies that share the spoils of victory, and find better places for these people in society?"

Survey participants concur that unemployment hype is over-baked; only 34% viewed reducing labor costs as an important business driver for deploying AI. Improving decision-making speed and quality, in contrast, was voted a top driver by a far higher share of 51%.

There is no "global"

It is intuitive to discuss AI ethics at a global level, but many research studies have shown that there are major cultural differences in what could be considered the "right" way to articulate ethics. "Ethics is linked with morality and culture, which differs from region to region and culture to culture,

"If we do it well, AI can improve our life dramatically, because we remove people from working in toxic or demeaning environments. The question is, what do we do with these people? Can we create societies that share the spoils of victory, and find better places for these people in society?"

Loredana Padurean
Associate Dean
Asia School of Business

and even within a single country," says Zee Kin Yeong at IMDA. Cultural factors shape how citizens view the penetration of AI into realms like the home. 'Care-bots' for providing home-based health care and social services have taken root in Japan, where robots are ascribed human qualities. "People [in Japan] think about AI not as a computing machine, but more like a human being," says Professor Kenji Suzuki, Center for Cybernics Research and Faculty of Engineering at the **University of Tsukuba**. "Different attitudes to AI lead to different approaches to developing it within society."

The infamous Moral Machine trolley experiment posits that a car's brakes fail, leaving it to either stay on course and kill three elderly people—two men and a woman—who disobeyed a "do not cross" signal, or swerve and kill its three passengers: an adult man, a woman, and a boy. The Moral Machine project analyzed 40m decisions from 233 countries

to reveal that respondents from collectivist cultures, like China and Japan, were less likely to spare the young over the old.[18] The element of risk-taking, that some pedestrians were breaking the rules, also played a part in the decision-making of the respondents in some countries.

Unintended consequences

A third critique is that of unintended consequences from proposed regulatory responses. Algorithmic transparency, for instance, seeks to improve accountability, transparency, and fairness by mandating the sharing of information about the functioning of algorithms responsible for automated decisions. It is already outlined in the European Union's General Data Protection Regulation and elsewhere, including France's Digital Republic Bill. Yet some tech companies and technology industry groups warn that such approaches could, by revealing how algorithms work, enabling hacking, the "gaming" of algorithmic systems, and even intellectual property theft.[19] It is also unclear how the working of complex algorithms can be translated into language that the unskilled can meaningfully understand.

Rather than developing new AI regulations or codes, it might be more practical and achievable to ensure that AI does not transgress existing civil rights. Some legal scholars argue that AI codes are ambiguous and lack accountability. More effective would be to govern AI according to existing institutions, like the international human bill of rights; if an AI removes a person's rights, it should not be acceptable.[20]

While some AI innovations might be genuinely new (such as autonomous vehicles), experts believe we are still very far from the dawn of general AI, which would genuinely challenge our existing rules. "We are very far still from really building or knowing how to build intelligent machines—intelligent in a sense of not just better at playing chess, or golf, or driving a car," says Tomaso Poggio at **MIT**. "This is the greatest problem in science, and could be up to 50 years away." There are many other "more pressing and immediate risks to mankind, from global warming to existing nuclear weapons."

Key takeaway

A global approach is unlikely—Many countries and international organizations are forming panels and working groups to understand the risks posed by AI and create frameworks for guiding its future development. The emergence of a single framework is unlikely given the many subtle differences around the world in culture and perception of ethical rights and wrongs. Yet the most important thing is interoperability, that the frameworks being developed are flexible so that the ecosystem can thrive.

Guidelines for growth—Executives in Asia understand that AI must be developed in a way that is responsible and transparent. Communicating clearly with customers and stakeholders about how data is gathered, used and shared will continue building trust. Innovative policy approaches around data governance are also likely to emerge.

[18] *www.nature.com/articles/s41586-018-0637-6*
[19] *www.techcrunch.com/2018/03/12/report-calls-for-algorithmic-transparency-and-education-to-fight-fake-news/?guccounter=1*
[20] *www.technologyreview.com/s/612318/establishing-an-ai-code-of-ethics-will-be-harder-than-people-think/*

7. Conclusion

In Asia, AI is being industrialized and commercialized at scale, humming in the background everywhere from hospital diagnostic clinics to smartphones, banks, and military facilities. Concern about the potential threats to societies and individuals that this technology poses may be outweighed by confidence about the benefits, but it is not absent. This report's conclusions are:

1. **Frameworks over regulation.** Stakeholders in Asia's AI ecosystem are calling for greater pragmatism and precision in the way that the AI ethics debate is framed. Some risks are already covered by regulations, whether human rights laws or data privacy rules. Survey respondents are looking for governments to lead, but any policy, frameworks, or regulation must be created sensitively and without stifling innovation.

2. **Transparency and trust.** Governments can focus on building a trusted and socially beneficial AI ecosystem by strengthening the foundations for AI development. This includes ensuring that data is collected in an ethical way and holding companies and organizations to account for bias or discriminatory practices. Trust-building also involves public education about AI and managing data controls and privacy. Proponents of algorithmic transparency argue that data processes and the ways in which AIs make decisions should be open and accountable, preventing the erosion of human rights.

3. **The best and the worst of AI.** In the decade ahead, it is likely that Asia will see some of the fastest applications of AI in areas such as healthcare, agriculture, financial services, and many other areas that benefit people's lives. Already, countries like China are racing ahead in the use of AI to democratize access to medical services. Yet the flip side is that Asia is also already witnessing some of the most invasive and controversial uses of AI for state surveillance and social control. This is likely to continue; the establishment of supranational ethical principles guiding the use of AI by governments will be a vital step.

4. **Maintaining harmony between humans and machines.** Our survey shows that business leaders are concerned that AI will remove more jobs than it creates across the region. Private and public sector organizations must think broadly about people capabilities and invest heavily in reskilling. The businesses that are able to, for example, train displaced finance staff to become IT and cybersecurity specialists will save substantially on recruiting costs in the long term. For policymakers, the challenge is to balance the continual push toward industrial digitalization with investments in long-term human capital development.

About MIT Technology Review Insights

For more than 100 years MIT Technology Review has served as the world's longest-running technology magazine, the standard bearer of news and insights on how the latest technologies affect the world around us. Read by a global community of innovators, entrepreneurs, investors and executives at the highest level, it offers an unrivaled authority that is backed by the world's foremost technology institution, and features editors with a deep technical knowledge and understanding of technological advances.

MIT Technology Review Insights is the content solutions division of MIT Technology Review. It includes two main divisions: Research and Live Events. Aligned with the same stellar editorial heritage and standards as the magazine itself, we leverage our access to a wide network of subject matter experts and leading content contributors to create custom content for clients who want to reach new audiences with relevant, cogent and high-quality stories and experiences to users wherever they want it—in digital, print, online, and via unique in-person experiences.

MIT Technology Review Insights

insights.techreview.com

@techreview @mittr_insights

insights@technologyreview.com

Is the AI ethics issue hindering innovation?

Source: Shutterstock

By **Mark Jones** | 4 February, 2021

- **There are too many cases of discrimination in AI and, in some cases, organizations are even shelving plans to adopt it altogether**

Beneath the exciting tide of artificial intelligence (AI) applications permeating industries and consumers' daily lives, there has been an undercurrent growing in strength over years: the question over whether we can trust the decisions of morally-void autonomous systems, informed by and interpreting only the datasets they receive.

The challenges around ethical AI have, for several years, been viewed as the biggest challenge facing its users, but now some organizations are actually killing plans for adoption because of the potential danger or avoiding embarking on projects altogether.

As reported by *Yahoo*, the chief technologist for business software maker LivePerson, Alex Spinelli, said he had canceled some AI projects at his employer (and previous ones) over concerns about AI – in particular, the use of machine learning to analyze customer data and make predictions about behavior.

Spinelli attributed the AI systems – particularly those used by Facebook to target users with content and pages it thinks will be of interest to users – to the spread of disinformation conflagrating the pro-Trump Capitol riots last month.

AI has shown transformative potential in its ability to undertake complex tasks with lower costs and resources. Use cases are proliferating, from detecting fraud, increasing sales, improving customer experience, automating routine tasks, to providing predictive analytics, while automated chatbots remain the most widely-adopted machine learning

CLOUD DATA CENTER CYBERSECURITY 4G
ZERO-TRUST OPEN SOURCE AUDIO-VISUAL
INDIA CHIPMAKER DATA BREACH

application.

But when it comes to the question of AI ethics, there are plenty of examples to show we're a long way off the mark.

Bias in AI-powered facial recognition systems is perhaps the most prolific example. In 2018, a study by MIT found that while determining gender using three different facial recognition programs, the error rate for light-skinned men was 0.8%, while darker-skinned women were misgendered 20% to 34% of the time.

Amid Black Lives Matter protests last year – and a flurry nationwide of what was deemed excessive force by the law – IBM withdrew its facial recognition technology, condemning the wider technology's use for "for mass surveillance, racial profiling, violations of basic human rights and freedoms."

Amazon then also withdrew its Rekognition software from use by law enforcement. But the company was earlier forced to park its AI candidate-screening technology due to an inherited lack of gender-neutrality. The 'secret' tool was supposed to rank candidates with a five-star rating system. Amazon previously canned another AI-powered recruitment program, after discovering that the 10 years' worth of successful applications it was consulting to make decisions were male-dominated, and therefore unfavorably discounting women.

Most recently, Google's former co-head of AI ethics and a prominent black female researcher at the company, Timnit Gebru claimed to have been fired after the company blocked the publication of a report she co-authored, raising ethical questions around the use of large, data-consuming language models in which Google is one of the leaders. Following her departure, the search giant went on to suspend computer access to another of the firm's AI ethics researchers who had been critical of the company.

A need for regulation

With machine learning models relying on algorithms learning patterns from vast pools of data, models are at risk of perpetuating bias present in the information they are fed. AI's mimicking of real-world, human decisions is both a strength and a great weakness for the technology— it's only as 'good' as the information it accesses. Of course, this challenge isn't new; as innovation continues, AI and machine learning ethics are regularly touted as crucial to the technology's development. This challenge is on the radar of organizations, world governments, and the machine learning community. To date, there has been a growing body of work on ethical AI principles, guidelines, and standards across different organizations, including IEEE, ISO, and the Partnership on AI. But guidelines are still lacking, and many organizations are navigating the complex waters of self-governance.

In 2019, a Vanson Bourne study revealed 89% of IT heads believe AI development should be regulated, with the need for a level of control and central oversight deemed necessary, even if it hindered the pace of the technology's evolution and applications by organizations.

Self-regulation and governance – and the creation of internal AI ethics panels – aren't keeping pace with AI's growing scale and sophistication. A report by Pegasystems found that 65% of respondents felt current external governance was insufficient

to manage AI adoption; 70% of respondents expressed fear about AI.

Effective self-governance requires enterprises to check software for AI algorithms are correct, and that the algorithms are ethical. But despite AI's proliferation, just 27% of respondents have a designated leader in AI governance, with Manufacturing, Healthcare, and Financial Services all reporting significant gaps in internal leadership and formal strategies.

There is **plenty of advice** out there for organizations to stay on the ethical path with AI, but many organizations seem to be crying out for more hardline guidelines, consistent across industries.

As far back as 2017, Elon Musk called for the regulation of AI development, despite being "clearly not thrilled" to be advocating for government scrutiny that could impact his own industry. The Tesla CEO believed the risks of going without were simply too high.

"Normally the way regulations are set up is when a bunch of bad things happen, there's a public outcry, and after many years a regulatory agency is set up to regulate that industry. It takes forever," he told *NPR*.

"That, in the past, has been bad but not something which represented a fundamental risk to the existence of civilization."

So far in the US (and preceding the new administration), there's been an appetite for "light touch" regulation built of flexible frameworks, with the intent to do as little as possible to stymy the growth of the country's technology industry. Under the Trump administration, the Office of Science and Technology Policy warned against more hardline policies being nudged forward in Europe: "Europe and our allies should avoid heavy-handed innovation-killing models, and instead consider a similar regulatory approach."

In an article entitled *AI That Reflects American Values*, United States' chief technology officer Michael Kratsios wrote; "The U.S. will continue to advance AI innovation based on American values, in stark contrast to authoritarian governments that have no qualms about supporting and enabling companies to deploy technology that undermines individual liberty and basic human rights.

"The best way to counter this dystopian approach is to make sure America and our allies remain the top global hubs of AI innovation."

In that same Pegasystems survey, however, concerns were not so much that regulation would dampen or hamper innovation or make adoption more complex and expensive – they leaned towards a worry that regulation would be insufficient to manage AI adoption.

With no universal regulation set-in-stone, businesses are left to make their own assessments on AI and how to ensure the way it's applied is deemed ethical. They must consider whether the business benefits they gain from the technology are worth the risk of discrimination.

The fact that the question of AI

ethics is giving some business leaders pause is a good sign, at least, and some organizations are notably making progress themselves.

THE QUESTION MARK OVER AI AND INTELLECTUAL PROPERTY

Mark Jones | 30 September, 2020

Last year, engineering giant Rolls-Royce unveiled a "workable, peer-reviewed AI ethics framework" published under Creative Commons license, which it said can help any organization ensure the decisions it takes to use AI in critical and non-critical applications are ethical. The company has been using AI for around two decades, including using the technology to monitor jet engines in service in real-time. But as it looked to extend AI to more parts of the business such as robotic inspections of critical components, it was becoming critical to address rising concerns around ethical and transparent AI.

"Rolls-Royce's AI capabilities are embedded deeply into other companies' products and services and so aren't widely known. Rolls-Royce's AI doesn't often feature in a consumers' understanding of how the digital world is changing their lives," said Caroline Gorski, Global Director of R2 Data Labs.

"The current debate about the use of AI is focused on the consumer and the treatment of consumer and personal data. But we believe that what we have created – by dealing with a challenge rooted squarely in the industrial application of AI – will help not only with the application of AI in other industries but far more widely," she added.

Rolls-Royce chief executive, Warren East, stressed that the firm wants to move from "conversation" around concepts and guidelines of AI ethics to "applying it".

"There is no practical reason why trust in AI cannot be created now. And it's only with the acceptance and permission of our society – based on that trust – that the full benefits of AI can be realized, and it can take its place as a partner in our lives and work."

ARTIFICIAL INTELLIGENCE | AUTOMATION | BIG DATA | MACHINE LEARNING

Mark Jones | @mjtechno

Mark is Editor at Tech Wire Asia and TechHQ (part of the Hybrid group).

READ MORE

AWS has invested US$6.5 billion in Singapore since the launch of its cloud region in 2010

Equinix invests US$74 million to build data center in Indonesia

One-size-fits-all is a myth when securing cyber-physical systems

Can Maxis' metaverse experience live up to customers' expectations?

Dell hoping to simplify zero trust for organizations

Opinion

Michael Kratsios

AI That Reflects American Values

We don’t have to decide between freedom and technology.

Friendly robots... nice robots... *Photographer: Robyn Beck/AFP/Getty Images*

By Michael Kratsios
January 7, 2020, 1:58 PM UTC

Innovations in artificial intelligence are creating personalized cancer treatments, improving search and rescue disaster response, making our roadways safer with automated vehicles, and have the potential for so much more.

But with growing concerns about data privacy, big tech companies, and the rise of technology-enabled authoritarianism in China and elsewhere, more people are starting to wonder: Must we decide between embracing this emerging technology and following our moral compass?

That's a false choice. We can advance emerging technology in a way that reflects our values of freedom, human rights and respect for human dignity.

As part of the Trump Administration's national AI strategy—the American AI Initiative—the White House is today proposing a first-of-its-kind set of regulatory principles to govern AI development in the private sector. Guided by these principles, innovators and government officials will ensure that as the United States embraces AI we also address the challenging technical and ethical questions that AI can create.

These regulatory principles are designed to achieve three goals: Ensure public engagement, limit regulatory overreach and promote trustworthy technology.

First, we're encouraging federal agencies to provide opportunities for public comment in AI rulemaking, including feedback from the American public, the academic community, industry leaders, non-profits and civil society.

The principles also promote a light-touch regulatory approach. The White House is directing federal agencies to avoid preemptive, burdensome or duplicative rules that would needlessly hamper AI innovation and growth. Agencies will be required to conduct risk assessments and cost-benefit analyses prior to regulatory action to evaluate the potential tradeoffs of regulating a given AI technology. Given the pace at which AI will continue to evolve, agencies will need to establish flexible frameworks that allow for rapid change and updates across sectors, rather than one-size-fits-all regulations. Automated vehicles, drones, and AI-powered medical devices all call for vastly different regulatory considerations.

Finally, our new regulatory principles promote the development of trustworthy AI. When considering action related to AI, regulators must consider fairness, transparency, safety, and security. Agencies should also pursue verifiable, objective evidence for their policy decisions, basing technical and policy decisions on the best possible scientific evidence.

The White House calls on agencies to protect privacy and promote civil rights, civil liberties, and American values in the regulatory approach to AI. Among other important steps, agencies should examine whether the outcomes and decisions of an AI application could result in unlawful discrimination, consider appropriate measures to disclose when AI is in use, and consider what controls are needed to ensure the confidentiality and integrity of the information processed, stored and transmitted in an AI system.

The U.S. will continue to advance AI innovation based on American values, in stark contrast to authoritarian governments that have no qualms about supporting and enabling companies to deploy technology that undermines individual liberty and basic human rights.

Governments elsewhere are co-opting companies and deploying their AI technology in the service of the surveillance state, where they monitor and imprison dissidents, activists and minorities, such as Beijing's treatment of the Muslim Uyghurs.

The best way to counter this dystopian approach is to make sure America and our allies remain the top global hubs of AI innovation. Europe and our other international partners should adopt similar regulatory principles that embrace and shape innovation, and do so in a manner consistent with the principles we all hold dear.

This would build on the success we have already found in working with our international partners. The Trump Administration has enhanced cooperation on AI research and development, and we committed at the G7 to work with our allies on how AI can be applied to solve challenges that reach across borders. With our allies in the OECD, we reached consensus on international principles to improve the trustworthy development of AI without holding back AI innovators.

Here in the United States, the White House has underscored its commitment to public engagement in AI by opening our regulatory principles for public comment. By working together, we will shape the policies that guide how AI is developed and deployed so that all people and communities can enjoy the benefits and opportunities it provides.

Americans have long embraced technology as a tool to improve people's lives. With artificial intelligence, we are ready to do it again.

Michael Kratsios is the Chief Technology Officer of the United States.

This column does not necessarily reflect the opinion of the editorial board or Bloomberg LP and its owners.

To contact the author of this story:
Michael Kratsios at press@ostp.eop.gov

To contact the editor responsible for this story:
Sarah Green Carmichael at sgreencarmic@bloomberg.net

The case for uniting the East and West to build ethical AI

AP PHOTO/ANDY WONG

A Chinese woman arranges the Chinese and U.S. flags on stage before both countries' delegates group photo session for Third China-US Strategic Economic Dialogue at…

FROM OUR SERIES
Ideas
Our home for bold arguments and big thinkers.

By **Soraj Hongladarom**
Director of the Center for Ethics of Science and Technology, Chulalongkorn University

Published May 21, 2019 • Last updated July 20, 2022 • This article is more than 2 years old.

In March 2016, the AI program AlphaGo shocked the world by

defeating the world champion Lee Sedol in a widely publicized event watched by more than 200 million people. Go, the ancient game most likely originated in China, was until then believed to be beyond the power of any machine. But in the end, the program developed by DeepMind—a British-based company funded by American Alphabet dollars—soundly defeated the legendary South Korean champion.

When it comes to AI, there is a tension between East and West. In AlphaGo's case, an ancient game of the East was pitted against a software program developed in the West. The defeat sparked a surge in interest in AI in many Asian countries, and now China is hot on the US's robotic heels when it comes to AI funding.

Society is witnessing a blossoming of cultural traditions underlying both the design of, and the expectations for, AI. This new confluence in ethical thinking will hopefully arise from this collective—one that is beneficial toward a common good. But that's not guaranteed.

Today's AIs are ubiquitous and contain the potential and power to change the world in ways we can scarcely imagine, for better or worse. This exponential surge in AI has thankfully prompted an interest in ethics and ethical considerations for it. However, much of this thinking is dominated by Western theories.

As the field has become more global, more cultures, particularly Eastern ones, are becoming more active players in the field. The thinking behind AI ethics therefore needs to include the traditional thinking systems of these cultures. Though China is close to dominating global business and research in the field, other cultural influences are finding a place too, including those from Africa, the Middle East, and the rest of the Global South.

We can no longer simply apply a Western value set to AI—but we shouldn't start applying a wholly Eastern set, either. As AI becomes more global, the theories that underpin its ethics must therefore also take on global dimensions

The difference between Western and Eastern philosophy

What are the main differences between Western and non-Western underpinnings in AI ethics?

We should start by acknowledging that both the West and the East (including Africa and the Middle East) are vast regions, each containing a huge variety of different traditions. But despite the numerous differences and expectations in our varied traditions, we find common ethical ground in the development and design of AI.

Western ethics are dominated by ethical theories such as deontology and utilitarianism. The first emphasizes the use of reason and logic to find what is believed to be the right answer to ethical problems, and then demands adherence to moral decisions irrespective of the consequences. The second seeks the greatest good for the greatest number of people. This seem to be a simple and practical solution, but the difficult part is working out what exactly is the good and how to measure its quantity.

> **“Honesty, truthfulness, compassion, and altruism are praised across all traditions, East or West.”**

Both the dominant theories are based on the individual as the determining factor in judging whether an action is good or bad; they’re individualistic in the sense that both reason and utilities belong to particular individuals. There is another theory—virtue ethics—which looks at cultivation of moral character as key to ethical judgment. Of the three, virtue ethics appears to be closest to the theories of the East.

The perspective is different when we consider the dominant traditions of ethics in the East, however. For example, the Ubuntu tradition in Africa advocates a focus not on the individual, but the community. The word "ubuntu" is a Bantu term meaning "humanity," or more precisely "the bond that binds all of humanity together in a single whole." This bond is given preference over single individuals—the difference is between humanity being put together as a single unit and atomic individual persons. Dominant Western theories emphasize the latter, with the result that it gives rise to individuals building walls against one another, both literally and metaphorically. With the Ubuntu concept, on the contrary, individuals seek out others, and hold hands together.

Buddhism also has a profound view on ethics. An ethical action is one that leads the one doing it to achieve the eventual goal, which is supreme happiness. This is the kind of happiness that results when one is completely attuned with nature: It is not the same as individual preferences or utilities as in utilitarianism. In other words, supreme happiness for Buddhism is not the same as supreme pleasure, and it is not individualistic because it is the same for everybody.

Furthermore, the indigenous spiritual system of Japan—shinto—holds that there are spirits everywhere, such as in forests, wind, water, and so on. Its ethics are thus based on harmony with nature; an unethical action is one that breaks the bond that humans already have with their natural environment.

Applying the similarities to AI

Nonetheless, even with these great differences among these traditions, we can still find some similarities. It is these intersections that paves the way toward a truly global ethics in the age of AI.

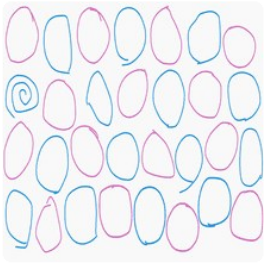
Honesty, truthfulness, compassion, and altruism are some of the examples that are praised across all traditions, East or West. One should be honest because doing so is to follow the universal maxim, according to the Kantian strand of Western ethics. On the other side of the world, the Ubuntu and Shinto traditions believe that one should also be honest, because being honest reinforces the bonds that tie us all together. As philosophers and computer scientists collectively and globally ponder the question of the "good life," this is the basis for a global AI ethics system.

In the end, we need manufacturers of AI to focus on augmenting design with values like honesty, loyalty, truthfulness, and altruism. If embedded correctly, these attributes can be manifested to end users.

The environment that surrounds AI needs to be immersed in human-centric concerns that are common to all people, not just those in the East or the West.

After all, AIs are manufactured by people, and its system design needs to reflect human virtues. A system consisting of an ethical awareness of the people involved in its creation and implementation will create a technological future that accounts for culturally varied virtues.

POPULAR STORIES



DIFFERENT STROKES

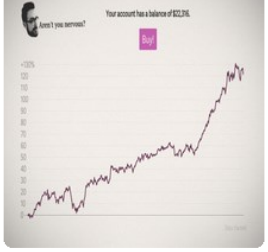
How do you draw a circle? We analyzed 100,000 drawings to show how culture shapes our instincts

Quartz • June 15, 2017

INDONESIA

How the Gujaratis took over Belgium's diamond trade

Quartz • October 17, 2014



GAME ON

This game will show you just how foolish it is to sell stocks right now

Quartz • August 26, 2015

WHAT DREAMS MAY COME

Why everyone around the world is having the same nightmare

Quartzy • October 30, 2018

NO MATH REQUIRED

BMI calculators aren’t accurate, but our body fat calculator is

Quartz • June 10, 2017

PAINT BY NUMBERS

The best examples of DALL-E 2’s strange, beautiful AI art

Quartz • June 10, 2022

NPM ERR!

How one programmer broke the internet by deleting a tiny piece of code

Quartz • March 27, 2016

NO BIGGIE

No, the US doesn’t have $30 trillion in debt

Quartz • February 2, 2022

DIPLOMATIC PLATE DECODER

Spot a US-issued diplomatic license plate? Decode it here.

Quartz • September 25, 2012

SPRING FORWARD FOREVER?

What will change if the US adopts daylight saving time permanently?

Quartz • March 17, 2022

EngageMedia: Alternative ethical frameworks for AI: A critical view of AI ethics

By Jun-E Tan for EngageMedia

Published on 31 August 2020
Page last updated on 23 September 2022

This article was rebublished from EngageMedia.

This article is the last in a series on the ethical principles and guidelines of artificial intelligence (AI), as well as their shortfalls and the search for alternative frameworks.

In Parts One and Two of this series on the ethics of artificial intelligence (AI), we dismantled the assumption that we can rely primarily on ethical guidelines for industry self-regulation, based on two perspectives: the substance of the ethical documents and the difficulties in putting principles into practice. The content of ethical documents were found to be narrow and limited in scope, while the implementation of beautifully crafted principles and guidelines proved to be difficult without the teeth of legal regulation.

Is that then the end for ethics in AI? That may too hasty of a conclusion. Luciano Floridi[1] points out that self-regulation cannot replace legal regulation, but it can still be a valuable tool for AI governance in situations such as when legislation is unavailable or when legislation is in need of an ethical interpretation. Sometimes, there are also situations when we need to make a judgment call if it is better to do (or not do) something, even if it is not illegal to do (or not do) it, such as improving labour conditions for workers in the gig economy even if it is not yet legally required.

Therefore, in this last part of a critical view of AI ethics, we will explore some studies and thought experiments of how ethical frameworks from other cultures and traditions can be brought in to scrutinise AI design, deployment, and use. The main point of this post is to explore existing alternatives of determining right and wrong, extending the discussion beyond what is available/used now in the tech industry in the West, focusing on the ways of thinking, and providing some examples of how they can be applied.

Challenging the universality of current AI ethical thinking

Scholars have drawn from diverse philosophies such as Confucianism (East Asian), ubuntu (African), or indigenous epistemologies to question the universality of the prevalent model of AI ethics and propose alternatives they consider to be better. In this blog post, we will not have enough of space to do justice to the individual philosophies, and can only pick and choose some points here and there to provide some ideas. However, just acknowledging that alternatives exist provides some level of empowerment – that it is not a choice between poor and nothing – but that we can take one step outside the box and look at other offerings, perhaps even from our own cultural knowledges and wisdoms in the context of culturally diverse Southeast Asia.

There are certain commonalities in the considerations of these alternative philosophies in contrast with what we have seen as current AI ethics in Part One, which is mainly drawn from a Western tradition of ethics. We will tackle three of these alternative philosophies as follows:

1. A different conception of "personhood", or what it means to be a person or a human
2. The importance of context, where ethical judgments are situated in their time and place, with nuances that may change the judgment
3. The importance of the relational aspect, between human and human, human and other beings, and human and AI

In the following sections we will look at these various points that help us understand some alternatives on how to do AI ethics better.

How to be good, how to be human

In Thilo Hagendorff's[2] critique of AI guidelines, he points out that there are different strands of ethics theory, and "deontology", which emphasises codes of conduct that lay out rules and duties (precisely our lists of ethical principles and guidelines), may not be the best route to take in terms of AI ethics. Instead, he points at another strand of ethics theory, "virtue ethics", which focuses instead on the moral character of the individual (developer, tech company, user, and more) and not the technology itself. Virtue ethics goes beyond ticking the boxes, and can be seen "as a project of advancing personalities, changing attitudes, strengthening responsibilities and gaining courage to refrain from certain actions which are deemed unethical" (pg. 112).

In other words, what would a good person do? And, further, what does it mean to be a good person or, indeed, a good human? Sabelo Mhlambi[3] states that the traditional Western view of personhood is based on rationality and that "truth could be rationally deduced through formal rules of logic" (pg.1). This philosophy is inherently individualistic (humanness as the individual's ability to arrive at the truth by logical deduction), and has motivated modern computing to build a machine that would match or surpass humans in reasoning or rational thinking. In terms of ethical behaviour, Mhlambi goes on to argue that the pursuit of rationality in such an interpretation has justified a host of dehumanising actions such as colonisation or racial subjugation of communities that are deemed not rational enough and hence not human enough. Endless economic growth and accumulation of capital also came part and parcel of what is considered as rational, leading to and justifying centuries of widening inequality that has now been carried forth to the digital era in the form of data colonisation and surveillance capitalism.

Watch this TEDx Talk by Getrude Matshe to learn more about the ubuntu philosophy.

A person is a person through other persons

In contrast, ubuntu, which is the basis of African philosophy, defines a person from the point of view of social relationships, as "fundamentally relational", where "a person is a person through other persons" (Mhlambi, 2020, p.3)[3]. From this departure point, ethics is considered from the point of view of one's relationality with other persons, their non-human counterparts, and the environment in general. One is recognised as a person or human only when they meet the responsibility of being humane to others, or when they improve the quality of the interconnected relationships that they are part of. Social progress is seen from the perspective of social harmony, which is in turn seen in terms of human dignity.

How would this be applied in the context of AI? Mhlambi critiques automated decision-making systems (ADMS) with the lens of ubuntu, and argues that ADMS can be flawed in five ways which violate the ethics of ubuntu:

1. When they exclude marginalised communities in their design
2. When they exacerbate current social and racial biases
3. When they fail to recognise the interconnectedness of society
4. When they commodify our digital selves, and
5. When they centralise data and resources in the hands of a few and enable them to inflict harm on the rest of society.

As we have explored in Parts One and Two, some of these aspects have completely been overlooked by the AI ethical principles that we have in place today, which argue that a system can be deemed ethical if it is fair, accountable, transparent (among other principles that focus on narrow fixes)– without taking into account the power dynamics and social relationships within the system.

Ethics as contextual, and a constant process of negotiation

Confucianism is another tradition that considers virtue in a relational way. Widely influential in East Asia, this philosophy originated from the teachings of Chinese philosopher Confucius (551-479BCE). Similar with ubuntu, Confucianism believes that humans are fundamentally interdependent, and one can only mature as a human in relation with *wulun*, the five types of social relationships that they are in: parent-child, sibling, husband-wife, ruler-minister, and friendship. One is not considered a proper human if they do not fulfil their duty and obligation within their social role. At the same time, they have to consider *dao*, or the right way to do things – dao of the heaven which is the principle that organises and governs the universe and/or the material world, and dao of the human which states that humans should live by acquiring virtues and cultivating morals.

The complexity of the relationships and ambiguity of considerations in Confucianism mirror life and its conflicting priorities and obligations. What if, for instance, what is right conflicts with what is good? Or what if self-interest conflicts with others' interest? Indeed, there are no easy answers to these questions, but I draw upon Pak-Hang Wong's[4] exploration of Confucianism and the ethics of technology for some ideas of how to address moral dilemmas. Primarily, we have to acknowledge that ethical behaviour is never a simple right or wrong, but is a constant deliberation about balance and harmony between different factors. While within Confucian ethics there are prescriptions of proper behaviour (e.g. on dao, and on social duties), it puts a lot of emphasis on "practising personhood", or "appropriately relating to and interacting with the others in various concrete situations, enabling a person to cultivate his or her moral sensitivity to the others and to the morally significant factors in the situation, which then allows him or her to comprehend relationships and situations more accurately and thus, to respond with propriety more effortlessly" (Lai, 2006, cf. pg. 78 of (Wong, 2012))[4].

The virtue ethics of Confucianism emphasises the moral character of an individual, and also the learning process of growing into this moral human who is able to interpret the situation based on its contextual factors and to decide on the best course of action given the circumstances. When technology joins the fray and changes societal equations, the Confucian does not seek a final answer to "solve ethics once and for all", but focus on balancing benefits and tradeoffs as a continuous process, through the lens of facilitating social roles and relationships.

Kinship or co-existence with AI

So far the ideas that we have discussed have focused mainly on AI and their societal impacts, but what about how we relate to AI? In a study on perspectives and approaches in AI ethics from an East Asian point of view, Danit Gal[5] suggests a tool-partner spectrum of how people in East Asia perceive AI, on one end as functional instruments to the other end as "friends, companions, romantic love interests, and fellow spiritual beings" (pg.2). Gal argues that the West mainly takes the view of AI and robotics as a tool, and while official government and corporate policies in China, Japan, and South Korea suggest the same direction, there is a divergence towards the partner perspective in academic thought, local practices, and popular culture.

Why is human-AI relationship important? Yi Zeng, a professor at the Chinese Academy of Science Institute of Automation argues that the safest approach to develop AI and robots is to give them a sense of self (consciousness) so that they would be able to empathise with human beings, and a reciprocal and respectful relationship will yield a beneficial outcome to both humans and AI (cited from (Gal, 2019))[5]. The Harmonious Artificial Intelligence Principles (HAIP) led by Zeng therefore does not only have a section on principles of AI respecting human rights, but also a section on how humans should treat AI, "including future conscious intelligent living becomings". Examples of the latter includes empathy ("What human do not want AI to do to human, human should not do unto AI") and privacy for AI. There is also a section of shared principles, where elements such as collaboration, coordination, mutual trust and evolvability are mentioned.

The same logic of a harmonious relationship between AI and humans has occurred elsewhere, in discussions of indigenous scholars who argue that indigenous epistemologies and cultural traditions of Hawai'i, Cree, and Lakota are able to provide conceptual frameworks that "conceive of our computational creations as kin and acknowledge our responsibility to find a place for them in our circle of relationships" (Lewis et al., 2018, pg. 4)[6]. They critique the Western "rationalist, neoliberal, and Christianity-infused assumptions" of AI which justifies the treatment of "the human-like" (referring not only to AI but also to indigenous communities which have historically been considered lesser humans by Western scientists and preachers) as slaves, while indigenous beliefs respect the animate and inanimate and their interconnectivities in a larger scheme of reciprocal and mutually beneficial relationships.

In conclusion

There is a plethora of alternative thinking beyond the current paradigm if we care to look. It is my hope

that this article has provided some windows to view AI ethics from various cultural landscapes to broaden the current discussion on how to make better and safer AI, and also to take a long-term view for when AI continues to develop from performing narrow functions to a general intelligence which emulates consciousness and independent thought.

Here I recall the "Good Way" of doing things in the native American Lakota tradition – it considers implications up to seven generations ahead. Indigenous scholars of the Lakota tribe are already exploring using indigenous protocols of ethical decision-making on ethical AI (Kite, 2020)[7].

How can a similar tradition be made applicable in the contexts of Southeast Asia and the Asia-Pacific, where there also exist diverse indigenous cultures? Any answer to this question is beyond the scope of this series, but if anything, the existence of alternative approaches to AI ethics only means that we can consider and define realities that make more sense to us and our respective contexts – those that are relational, inclusive, diverse, sustainable, and respectful to all parties involved.

For more information on AI in the context of Southeast Asia, check out this video, or go to Coconet.social/AI.

References

1. Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology*, *32*(2), 185–193. https://doi.org/10.1007/s13347-019-00354-x[↩]
2. Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8[↩]
3. Mhlambi, S. (2020). *From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance* (No. 2020–009; Carr Center Discussion Paper Series, p. 31). Harvard Kennedy School Carr Center for Human Rights Policy. https://carrcenter.hks.harvard.edu/files/cchr/files/ccdp_2020-009_sabelo...[↩][↩]
4. Wong, P.-H. (2012). Dao, Harmony and Personhood: Towards a Confucian Ethics of Technology. *Philosophy & Technology*, *25*(1), 67–86. https://doi.org/10.1007/s13347-011-0021-z[↩][↩]
5. Gal, D. (2019). *Perspectives and Approaches in AI Ethics: East Asia* (SSRN Scholarly Paper ID 3400816). Social Science Research Network. https://papers.ssrn.com/abstract=3400816[↩][↩]
6. Lewis, J. E., Arista, N., Pechawis, A., & Kite, S. (2018). Making Kin with the Machines. *Journal of Design and Science*. https://doi.org/10.21428/bfafd97b[↩]
7. Kite, S. (2020). How to build anything ethically. In J. E. Lewis (Ed.), *Indigenous Protocol and Artificial Intelligence Position Paper* (pp. 75–84). The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research (CIFAR). https://spectrum.library.concordia.ca/986506/7/Indigenous_Protocol_and_A...[↩]

Topic:
Artificial intelligence | Human rights and ICTs | Security and privacy | Digital society

Region: Asia

Tags: AI | ethics | information ethics | digital technology | artificial intelligence

« Go back

email@example.com Subscribe to our newsletter

ARTIFICIAL INTELLIGENCE POLICIES IN EAST ASIA:

AN OVERVIEW FROM THE CANADIAN PERSPECTIVE

ASIA PACIFIC FOUNDATION OF CANADA
FONDATION ASIE PACIFIQUE DU CANADA

TABLE OF CONTENTS

EXECUTIVE SUMMARY 3

INTRODUCTION 5

- Structure and Methodology 6
- Artificial Intelligence (AI) 101 6
- Why an AI Policy? 7
- Where is Canada in the AI Space? 9
- Why East Asian AI? 9

CHINA 11

- Government Policies 11
- AI R&D and Talent 13
- AI ELSI 14
- Conclusion 17

JAPAN 18

- Government Policies 18
- AI R&D and Talent 19
- AI ELSI 20
- Conclusion 22

SOUTH KOREA 24

- Government Policies 24
- AI R&D and Talent 27
- AI ELSI 28
- Conclusion 30

ANALYSIS & POLICY RECOMMENDATIONS 32

- Analysis 32
- Policy Recommendations 35

ACKNOWLEDGEMENTS 38

ABOUT THE ASIA PACIFIC FOUNDATION OF CANADA 39

EXECUTIVE SUMMARY

Accelerated advances in computing and greater integration of the internet – along with the proliferation of big data – have enabled significant progress in the field of artificial intelligence (AI) during the past few years, raising the profile of a discipline long relegated to the fringes. Many, especially in Asia, point to the DeepMind AI AlphaGo's victory over Lee Se-dol in the game of Go in 2016 as the watershed moment when the zeitgeist shifted. The possibilities – and therefore fears – tied to advances in AI have started to dominate the public discourse. Tangible progress in a field that blurs the line between science fiction and reality has helped fuel the furor and underscores the urgent need for intervention by policy-makers.

As AI is relatively new lay for people, it is challenging to talk about it, let alone create policies related to it. There are many unknowns and upcoming firsts in the regulation or governance of AI. In China, Japan, and South Korea, however, governments have begun creating AI policies in a comprehensive and committed manner. Thus East Asia provides excellent case studies for Canadian policy-makers as they begin to grapple with the development, deployment, and regulation of artificial intelligence.

Underpinning this report is the fact that all three East Asian nations regard AI as a key factor contributing to their international competitiveness, and that they have all taken nation-wide measures to enhance their AI capabilities. The scope of their national AI strategies is broad, with ambitious goals for everything from AI research and development (R&D) to integration of AI in specific economic sectors. This report provides a nuanced and detailed documentation of East Asia's AI agenda. It is hoped that the report will provide useful case studies for the Government of Canada as it seeks innovative ideas for expanding its own governance of this disruptive technology.

More importantly, this report asserts that the Government of Canada must recognize AI as a space for geopolitical competition and act proactively and strategically. The examination of the AI policies in East Asia paints a familiar picture for Canadians: China is a major influencer in the space of AI, promising opportunities for Canadians, but engagement with China entails risks and difficulties arising from differing values. In a similar vein, Japan and South Korea remain reliable partners for co-operation for Canadians, with the potential for collaboration yet to be fully exploited. Canada punches above its weight in AI research, and China, Japan, and South Korea are looking to Canada to help boost their own talent and research capabilities. Today, AI is a space where Canada can exercise greater influence, with the potential to take a leadership role in shaping norms of international AI governance.

Canada should, therefore, leverage its strengths and engage with these three East Asian countries in a strategic manner. It should participate in discussions on AI ethics and governance to positively contribute to the shaping of international norms, while building partnerships of like-minded nations to gain leverage in promoting uses of AI that align with Canadian values. As the technological becomes intertwined with the political, the perception of China as a threat to the

liberal world order continues to gain traction in the AI space. Western liberal democracies have strengthened their rhetoric and policies to counter China's approach to AI, highlighting liberal values in their ethics guidelines and policies. At the same time, China continues with its AI agenda and seeks to raise its international profile by becoming a norm-setter. Canada could play the role of the traditional "middle power" in facilitating constructive conversations, bridging the gap between China and the West. Meanwhile, Canada has the opportunity to form alliances with like-minded, relevant countries like Japan and South Korea to further increase its profile as a major player in international governance of AI.

As of July 2019, the trade war between China and the United States continues. This is accompanied by the United States' attempts to check the rise of Chinese high-tech firms, as exemplified in the White House and Department of Commerce orders to limit Chinese company Huawei's access to its U.S. vendors. The trade war underscores the greater role that technologies such as AI play in foreign policy today and brings attention to the need for the Canadian government to approach AI governance on the global stage in a strategic and comprehensive manner.

In this context, this report explores the national AI policies of these three East Asian countries to inform Canadian policy-makers and AI practitioners of the challenges and opportunities of a field that will soon intersect with every aspect of our lives. The report first establishes the relevance of AI and AI policies in East Asia for Canadian policy-makers. Then, three subsequent sections investigate AI policies in China, Japan, and South Korea, focusing on the policies, talent management, and ethical, legal, and social implications (ELSI) of the technology. Finally, the last section synthesizes the analysis of AI development in East Asia and concludes with four broad recommendations:

Recommendation 1: Increase Canada's capacity to govern AI in a co-ordinated manner at different levels of government.

Recommendation 2: Leverage Canada's AI talent more effectively as part of "Brand Canada."

Recommendation 3: Proactively address emerging AI issues in trade and investment.

Recommendation 4: Identify Canada's place in emerging AI geopolitics and engage strategically.

INTRODUCTION

It is an exciting time to be in the field of artificial intelligence (AI). Recent research breakthroughs and the transition toward a data-based society and economy have brought forth a golden age for AI. Every day, the media continues to circulate stories – often sensationalized ones – about AI and how it will revolutionize the way we live.

However, contrary to public expectations, AI will not dramatically change the world. Instead, it will make existing products and services faster and more efficient. AI does not exist on its own; as it is used today, AI is a program that is applied in existing hardware or services to maximize their efficiency – more similar to electricity than the T-800 from the film Terminator. But this means that AI, omnipresent like electricity, will influence – and already has influenced – most of what we do today, and this will have policy implications. AI presents the world with challenges and opportunities.

Particularly for Canada, this is a pivotal moment. During the late 1980s and early 1990s, at a time when AI was considered an interesting but impractical field, Canada invested in the field. Some of its institutions are now home to leading AI academics who are thriving in the field and attracting investments from tech giants like Amazon and Google. In the mid-2010s, the federal government invested C$125M on a national AI strategy to further support research and attract talent, which contributed to bringing Canada to the top of the list for investors and companies seeking to leverage AI. However, the upcoming transformations brought by AI pose new challenges to all stakeholders: government, industry, and civil society.

China, Japan, and South Korea's approach to AI provides important insights at this critical juncture for Canadians. Their governments have committed to comprehensive AI strategies that stretch beyond supporting research and development (R&D). These strategies point to opportunities for Canada in both its bilateral relations with these countries and in multilateral settings where Canada could take a leading role in driving the discussion/agenda around AI governance. Further, their strategies underscore the permeability and potential impact of the technology, which prompts us to think about our own approach to AI.

The objective of this study is to provide an overview of Chinese, Japanese, and South Korean AI policies, demonstrating their degree of commitment to not only research, but also commercialization, deployment, and considerations of ethical, legal, and social implications (ELSI). This report also analyzes the differences between Canadian and East Asian policies, drawing out implications for Canada. In order to limit the scope of this study, policy discussions on privacy and data management laws are not included. This report represents an attempt to make sense of the new space and provide the foundation for more in-depth research in specific areas of AI governance.

This report adds to the existing literature by taking a more in-depth look at the national policies in the three countries, with the intent of finding Canadian implications. At the moment, there are

several reports that focus on domestic implications (e.g. future of work) within Canada, or non-Canadian reports that focus specifically on China and threats to security. This report considers the region as a whole and provides comparative overviews while developing a greater analytical perspective on Canada's policy toward East Asia.

Structure and Methodology

The report is based on qualitative analysis of policy documents, reports, and media coverage, supplemented by interviews with relevant academics, policy-makers, and other practitioners in Beijing, Seoul, and Tokyo. These interviews occurred primarily during the spring of 2018 and are complemented by additional conversations.

The first section explains the basics of AI policy. Case studies of China, Japan, and South Korea are then presented. Findings are analyzed in the final section, and policy recommendations are presented. Each case study features an analysis of the broader government policy on AI, the country's attempts to address AI R&D and talent training, and its engagement with AI ELSI through government policy. Based on the analysis of the three cases in East Asia, four policy recommendations for Canada are presented at the end.

AI 101

How do we define *intelligence*? Philosophers have dedicated substantial tomes to this task, but disagreement remains, which leads to even more questions about the nature of intelligence. The definition of *artificial* intelligence becomes even more challenging, particularly when the term is conflated with other closely associated technologies like robotics and 5G, although neither of these constitute an artificial intelligence.

There is no singular definition of AI. Even Stuart Russell and Peter Norvig's leading textbook on AI, *Artificial Intelligence: A Modern Approach*, eschews a singular definition, and instead lists four different definitions: thinking humanly, thinking rationally, acting humanly, and acting rationally.[1] Russell and Norvig also note the interdisciplinary aspect of AI as a field, pointing to philosophy, mathematics, economics, neuroscience, psychology, computer engineering, control theory and cybernetics, and linguistics as the intellectual foundations of the technology, which underscores the point that defining AI could be tantamount to tackling the perennial questions in the discipline of philosophy.[2] Ryan Calo states that AI is "an umbrella term, comprised by many different techniques," and other policy documents on the subject take similar approaches.[3]

For the purposes of this report, AI is simply defined as simulation of human intelligence through computers, mainly referring to machine learning. Put simply, machine learning is a form of data

[1] Russell, Stuart and Peter Norvig. 2010. Artificial intelligence: A modern approach. Upper Saddle River: Prentice Hall, p. 2.

[2] Ibid., p. 5-16.

[3] Calo, Ryan. 2017. Artificial intelligence policy: A primer and roadmap. U.C. Davis Law Review. 51(2): 5.

analysis in which systems identify patterns and make decisions based on data in an automated manner. The theory itself is not new, but it took off in recent years due to the availability of big data and computing hardware capable of running these systems.

It should also be noted that we are still very far from seeing an AI like the T-800 from the Terminator movie franchise or Andrew in Bicentennial Man because current AI programs are limited to performing specific tasks that they were designed for – that is, they are "narrow" AIs. For instance, consider an AI program designed to play chess. A human may not be able to beat this program in chess, but this program will not be able to perform other tasks that we perform daily, such as figuring out the best transit route to work or summarizing the contents of a newspaper article.

The opposite of narrow AI is artificial general intelligence, a system that is able to perform *any* task that a human being can. But experts in the field generally believe that this will not happen in the near future. Martin Ford, an author who has written on AI and robotics, asked 23 leading experts when we should expect to see the first artificial general intelligence, and the average timeline of the 18 who responded is 2099.[4] As such, at least for the near future, the use of AI will mainly *complement*, not *displace*, humans or their intelligence.

This report focuses on the more immediate challenges emerging from greater integration of narrow AI into all aspects of our daily lives rather than issues surrounding artificial general intelligence.

Why an AI Policy?

If we are not concerned about the rise of a self-aware robot army, then why should we worry about AI at the policy level? It is because AI is a disruptive technology that will significantly affect the way businesses operate, as well as the larger society. According to a 2017 PwC report, AI is expected to provide a US$15.7T boost to global gross domestic product (GDP) by 2030, which would be equivalent to adding approximately 10 Canadian economies to the global economy.[5] Especially for high-income, developed countries with technical capabilities, AI promises an opportunity for increased productivity. Today, there is consensus around the world that policy-makers should support their respective AI research and industry to take advantage of the economic opportunity that AI presents.

However, this report argues that investment in AI policies is required not only for economic reasons, but also because of three broader challenges in the rapidly evolving field of AI.

First, most countries lack the necessary infrastructure to further their AI ambitions. There is a worldwide AI talent shortage, and major players in the private sector are racing to hire AI talent, from Silicon Valley to Beijing. According to a Tencent report, there are perhaps 300,000 individuals with AI expertise, while there is an immediate shortage of 800,000 specialists in this field.[6] This challenge will require a government-level response in education and immigration policies. The problem of infrastructure also extends to data, the "fuel" of AI. The greater the amount of data available, the better AI can be engaged. Today, however, most countries have regulations that

[4] Vincent, James. 2018. This is when AI's top researchers think artificial general intelligence will be achieved. The Verge. https://www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book; and Ford, M. 2018. Architects of intelligence: The truth about AI from the people building it. Birmingham: Packt Publishing.

[5] PricewaterhouseCoopers. 2017. Sizing the prize: What's the real value of AI for your business and how can you capitalise? Boston: PricewaterhouseCoopers.

[6] Tencent Research Institute. 2017. 全球 人工智能人才 白皮书 (Global artificial intelligence talent white paper). Beijing: Tencent Research Institute.

restrict data flow due to privacy concerns. As such, there is a need to reform existing regulations to allow the use of data, while ensuring the privacy rights of individual citizens.

Just as governments needed to create highways and establish rules of the road with the advent of the mass-produced automobile, so too do citizens and industry require the involvement of government to set up the basic infrastructure for an AI-driven economy to take full advantage of the new technology.

Second, AI involves ethical, legal, and social implications that will permeate different sectors. The best publicized example is labour market disruption due to automation. According to McKinsey, as much as 30 percent of human labour could be displaced by AI by 2030.[7] Further, AI involves potential threats to privacy rights, as well as the perpetuation and exacerbation of existing socio-economic biases. Economist Joseph Stiglitz warns that AI could further deepen the gap between the rich and the poor.[8] Experts in the field have raised concerns about the abuse of AI for malicious uses, from new forms of scams to drone attacks.[9] These concerns underscore the need for greater state involvement in proactively adjusting regulations to promote R&D and deployment, but also to prepare citizens for a society where AI is fully integrated.

Finally, AI raises potential challenges for state sovereignty. At the most basic level, integration of AI into weapons and weapon systems could change the nature of warfare. Russian President Vladimir Putin recently said that "whoever leads on AI will rule the world." The policy debates surrounding the technology are now increasingly framed within the language of an AI race, which emphasizes the competition between nations for supremacy on AI technologies, akin to the arms race between the United States and the former Soviet Union during the Cold War.[10] At a more subtle level, AI could also challenge the ability of states to govern effectively. AI, due to its decentralized nature, is a difficult technology to regulate. It has the potential to further undermine states through ELSI, such as mass unemployment or the dissemination of misinformation. As such, AI poses threats to a state's sovereignty at different levels, which warrants a response from policy-makers.

---

[7] McKinsey & Company. 2017. Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages. New York City: McKinsey & Company. https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages.

[8] Korinek, Anton, and Joseph E. Stiglitz. 2018. Artificial intelligence and its implications for income distribution and unemployment. NBER Working Paper No. 24174. Cambridge: National Bureau of Economic Research.

[9] Future of Humanity Institute. 2018. The malicious use of artificial intelligence: forecasting, prevention, and mitigation. Oxford: Future of Humanity Institute.

[10] Suder, Katrin. 2018. AI can change the balance of power. Berlin Policy Journal. https://berlinpolicyjournal.com/ai-can-change-the-balance-of-power/.

Policy-makers have become aware of the opportunities and challenges that come with AI, and governments around the world have started announcing their AI strategies.

Where is Canada in the AI Space?

Canada was an early investor in AI research. During the so-called "AI winter," a period during the 1980s and 1990s when decreased interest in AI led to a drop in research funding, Canadian funding agencies continued to support basic AI research, attracting scholars from all over the world. As a result, Canada today is home to world-leading researchers and their teams. At a time when AI talent is in high demand, Canada's research capabilities allow it to punch above its weight.

To further leverage this advantage, the Government of Canada announced in 2017 that it would commit C$125M for the next five years to further Canada's AI R&D and attract top-level talent in the country through the Pan-Canadian Artificial Intelligence Strategy, managed by the Canadian Institute for Advanced Research (CIFAR). The Strategy mainly addresses R&D and talent issues.

Further, there are various initiatives within the government that seek to address the advent of AI. Innovation, Science and Economic Development Canada (ISED) launched the National Digital and Data Consultations to understand the Canadian public perception of data issues. The Standards Council of Canada is working to develop a national standard for ethical uses of AI as well. The Treasury Board of Canada has also been active in the AI space. In February 2018, Canada joined the Digital Seven, a network of governments seeking to further integrate digital technologies in governance, and in July 2018 it created the first-ever Minister of Digital Government. In December 2018, the government also announced that it would collaborate with the French government on the International Panel on Artificial Intelligence, which will convene international AI initiatives.

Why East Asian AI?

There are three key reasons why Canadians should pay attention to East Asia's AI development.

First, China, Japan, and South Korea offer examples of comprehensive and forward-looking AI strategies that could provide lessons for Canadian policy-makers. The three East Asian governments have each formulated comprehensive national plans that promote AI R&D through facilitation of collaboration among policy-makers, academics, and industry stakeholders. Their plans also include strategic integration of AI into their economies and societies with a long-term view.

Second, understanding East Asian AI development is crucial for Canada's future engagement in the region. Economically and politically, the three East Asian countries are important partners for Canada. China is Canada's second largest trading partner. Japan and South Korea are also key trading partners (fifth and seventh largest, respectively). They are also regarded as vital partners in Canada's multilateral dealings and share similar political values. The national policies in East Asian states demonstrate a clear commitment to transforming their economies and societies through AI, which gives rise to new challenges in trade and diplomacy. Canada's continued engagement in Asia should be accompanied by a deeper understanding of the region's AI development.

Third, East Asian AI development underscores the emergence of an AI race and the need for Canada to identify its appropriate role. An examination of AI development in East Asia suggests a global fracture in attitudes toward AI and the subsequent emergence of an AI race or AI geopolitics. The attitude toward AI in East Asia clearly demonstrates that AI development is regarded as a national-

level project with implications in multiple areas of society. China, Japan, and South Korea regard AI development as a continuation of national development.

Combined with the cultural, social, and political factors that contribute to AI development, the consideration of AI as a national endeavour strongly suggests the emergence of AI as a space in foreign policy that could deepen existing splits in contemporary geopolitics. For instance, competition for securing AI talent and infrastructure between China and the United States has already been described as a duopoly – a race between the two – especially as China's use of AI for social surveillance and military application has alarmed liberal, Western states, led by the United States. Also, AI experts and practitioners in Japan and South Korea express concerns about their role "stuck between" China and the United States, sentiments that echo existing anxiety about China's rise in the region.

Canada has research capacities that allow it to perform well in the space of AI. Early investment in basic research has produced world-leading researchers in AI based in institutes at the University of Alberta, University of Toronto, and Université de Montréal, and has drawn investment from global giants such as Google and Amazon to the country. The Government of Canada launched the Pan-Canadian Artificial Intelligence Strategy in 2017 to further develop its existing strengths in research and talent. As the technological becomes increasingly intertwined with the political, these strengths in research and talent provide Canada with a tool to broaden its options in foreign policy, which should be leveraged in a timely manner.

CHINA

Government Policies

The Chinese Communist Party has clearly articulated a concrete vision for the role of AI in its policy documents released since 2015. The latest Five Year Plan (2016-2020), which is China's principal, large economic strategy, released in 2015, articulated China's intent to invest and gain leadership in big data and "intelligent manufacturing." Made in China 2025 consistently emphasized the need for China to upgrade its manufacturing sector by integrating emerging technologies (principally AI). As such, the Party had clearly identified the need to invest in AI for continued economic growth before the release of its AI-specific plans.

In July 2017, the State Council (the equivalent of the federal cabinet) released the Next Generation AI Development Plan, a comprehensive strategy document that outlines China's ambition to become a world leader in both AI development and application by 2030 (see Table 1). The Plan is comprehensive in its scope, addressing issues of talent acquisition, funding allocation, safety frameworks, technology standardization, and ELSI.

Table 1: Short and long-term objectives of China's AI development

Year	Goals	Scale of Core AI Industry	Scale of Related Industries
2020	Catch up to other nations in terms of AI research and application	RMB150B (C$30.5B)	RMB1T (C$203.3B)
2025	Achieve major breakthroughs academically and be world-leading in application	RMB400B (C$81.3B)	RMB5T (C$1.016T)
2030	Become the world's primary AI innovation centre	RMB1T (C$203.3B)	RMB10T (C$2.032T)

The AI Development Plan operates on four guiding principles.[11] First, it should be led by technology, in the sense that the policy-makers should be aware of the need to attain leadership in cutting-edge AI technology. Second, the Plan should be implemented systematically, placing basic research, technological R&D, industrial development, and commercial applications within a single system. It is noted that the "advantages of the socialist system" should be used to support this systematic approach to AI development, which points to greater policy latitude available to the central government. Third, the implementation of the Plan should be driven by the market, leveraging

[11] The author used the translation of the Plan posted in: China Copyright and Media. 2017. A Next Generation Artificial Intelligence Development Plan [translation by New America]. China Copyright and Media. https://chinacopyrightandmedia.wordpress.com/2017/07/20/a-next-generation-artificial-intelligence-development-plan/.

the resources in the private sector. It highlights the need to fully delineate the responsibilities of the government and the private sector to maximize the government's role in planning and guidance, policy support, security, regulation, environmental protection, and formulation of ethical guidelines. Finally, the Plan should follow the concept of open-source sharing and promote collaboration between industry, academia, research, and production units. It highlights the need to promote two-way conversion and application of AI technology for civil and military uses, and the need to continue participating in global research.

Following the release of the Next Generation AI Development Plan, the Ministry of Industry and Information Technology released the Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018-2020), which outlines more immediate guidelines for industry, government, and other stakeholders until 2020. The Action Plan highlights the importance of supporting research and entrepreneurship in establishing a foundational AI industry, and specifies target AI products such as autonomous vehicles, medical imaging diagnostic systems, video and voice identification systems, smart home products, and intelligent translation systems. The Action Plan falls in line with and supports the Next Generation AI Development Plan within a narrower time frame.

It is necessary to approach these national policies in a nuanced manner that considers China's unique form of governance. In the West, where there is emphasis on the rule of law, the text of a policy is of utmost importance, and any kind of adjustment in policy requires appropriate process (i.e. legislative procedures). However, this is slightly different in China. The text of a policy is important to some degree, but policies and regulations can be bent following mutual understanding between officials that such action is necessary to achieve greater goals. This can provide greater agility and flexibility in addressing emerging technologies such as AI (this advantage comes with the cost of lack of democratic accountability or predictability, of course).

In this case, the greatest impact of the Next Generation AI Development Plan is the Party's explicit endorsement of the industry and the implicit understanding that it will be followed by policy support. Ultimately, the Plan functions not as a detailed blueprint whose success is measured by congruence to its details, but as a general "wish list," according to Matt Sheehan of MacroPolo, a University of Chicago think tank that analyzes the Chinese economy. Sheehan writes that this wish list signals the relevant players – local government officials, private firms, academics – to deliver the items on it using their resources and ingenuity and promises the necessary policy support.[12]

Since the launch of the Next Generation AI Development Plan, China has made great strides in this area by leveraging its unique policy strengths. In China, these national plans offer an opportunity for ambitious local government officials to make a name for themselves by delivering the items on the wish list. Fifteen out of 34 local governments had announced their own AI strategies as of October 2018. The projected size of the AI industry from the combination of the 15 AI strategies amounts to RMB429B (C$80.17B) by 2020, almost three times the national goal of RMB150B (C$30.5B), which underscores the impact of the national directive upon local policies.[13]

The private sector has followed suit as well. The Chinese startup scene is notoriously competitive, and the nod from the government was quickly understood by ambitious entrepreneurs as an

[12] Sheehan, Matt. 2018. How China's massive AI plan actually works. MacroPolo. https://macropolo.org/chinas-massive-ai-plan-actually-works/.

[13] Ding, Jeffrey. 2018. Latest Policies for AI Industry [Data sourced from Qianzhan Chanye Institute]. https://docs.google.com/document/d/1NQ8kjeKso6mpaQtXyoFu7Tbl44UapdnNt7dOjGwhB8w/edit#.

opportunity to profit. Chinese startups received 48 percent of total global AI funding in 2017.[14] SenseTime, a Beijing-based facial recognition firm, became the world's most valuable startup at US$4.5B, and its success would not have been possible without the government policy support.[15] Chinese AI research is gaining more influence in academia as well, with the number of papers from Chinese researchers accepted by the Association for the Advancement of Artificial Intelligence conferences (the most prestigious in the field) more than doubling – jumping from 10 percent of papers to 23 percent – between 2012 and 2017.[16]

The Party has also taken further measures to maximize the efficiency of the private sector by organizing a National AI Team. More specifically, the Ministry of Science and Technology has designated tech giants such as Baidu, Alibaba, Tencent, and iFlyTek to lead specific sectors of AI. For instance, Baidu is responsible for autonomous driving, Alibaba for smart cities, Tencent for health care, and iFlyTek for voice recognition. The expectation is that the members of the National AI Team develop "open innovation platforms" on these key areas and thereby set the standards for each industry.[17] Further, this has the effect of minimizing competition among large tech firms, which have shown the tendency to compete until they have edged each other out of the market.

AI R&D and Talent

The issue of AI talent is a major policy issue not only in China, but in the rest of the world. This is a murky area for examination, as it is challenging to define "AI talent." For the purpose of this report, AI talent does not merely refer to high-level AI scientists and engineers with graduate degrees, but also workers who are able to function in AI-integrated workplaces (such as smart factories).

That being said, the shortage of talent – as an inclusive term that stretches beyond PhD holders – has been defined as a major challenge in China's AI ambitions. This has been explicitly indicated in China's official AI policy documents, including the Next Generation AI Development Plan, which notes that "cutting-edge talent for AI is far from meeting demand."[18] Tencent has also noted that the AI talent shortage is a major policy issue and that there may be a deficit of up to five million qualified workers within a few years.[19]

---

[14] Snow, Jackie. 2018. China's AI startups scored more funding than America's last year. MIT Technology Review. https://www.technologyreview.com/the-download/610271/chinas-ai-startups-scored-more-funding-than-americas-last-year/.

[15] SenseTime has gained access to footage from China's 170 million closed-circuit TV cameras by partnering with government, which in turn uses the technology for social surveillance. See Russell, Jon. 2018. China's SenseTime, the world's highest-valued AI startup, closes $620M follow-on round. TechCrunch. https://techcrunch.com/2018/05/30/even-more-money-for-senstime-ai-china/.

[16] Ding, Jeffrey. 2018. Deciphering China's AI dream: The context, components, capabilities, and consequences of China's strategy to lead the world in AI. Oxford: Future of Humanity Institute.

[17] Goodrich, Jimmy and Paul Triolo. 2018. From riding a wave to full steam ahead. New America. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/riding-wave-full-steam-ahead/.

[18] See the translation of the Next Generation AI Development Plan: https://chinacopyrightandmedia.wordpress.com/2017/07/20/a-next-generation-artificial-intelligence-development-plan/.

[19] Tencent op. cit.

The Chinese government views addressing this talent shortage as a priority. Even before the launch of the AI Development Plan, the Chinese government had been proactive in recruiting international academic talent in science and engineering through the likes of the Thousand Talents Program, which focuses on repatriating Chinese researchers working abroad. Further, since the launch of the Plan, all relevant players involved in China's AI development have been racing to attract talent. For instance, the municipality of Beijing introduced an incentive program for top AI talent in March 2018 that offered RMB1M (C$187,000) in cash prizes and a fast-tracked immigration process as perks.[20]

The private sector plays a key role in this space. Companies such as Baidu, Alibaba, and Tencent have been leveraging their financial resources to compete against other tech giants in the Silicon Valley and elsewhere to attract AI talent from around the world. Nick Zhang, president of the Wuzhen Institute, said experienced AI researchers are being offered a salary of US$1M or more by major companies.[21] The combined efforts of the Chinese in this space seem to be yielding results. According to a Tencent report, the proportion of foreign AI talent in China increased from 5.2 percent to 7 percent between 2015 and 2017.[22] In addition, these companies have also established AI labs outside China, such as Baidu's driverless cars laboratory in the Silicon Valley or Tencent's AI research lab in Seattle.

However, in the long run, China's priority is to strengthen its own AI talent pool, and it has started investing in AI education at the national level. In April 2018, the Ministry of Education issued the AI Innovation Action Plan for Colleges and Universities, which seeks to make Chinese research and academic institutions the leaders of AI innovation and training by 2030, in line with the goals of the Next Generation AI Development Plan.[23] Under this Action Plan, 100 AI specialization programs for specific domains (e.g. AI plus health care, AI plus transportation) would be created by 2020 through the development of 50 sets of world-class teaching materials, 50 national-level, high-quality online courses, and 50 AI institutes. By doing this, the Chinese government addresses the talent shortage in not only academic research, but also in lower-level, technical applications in different sectors that seek to integrate AI.

Further, under this initiative, AI expert and venture capitalist Kai-Fu Lee's Sinovation Ventures, created in 2017 to offer AI talent in Beijing, partnered with the Ministry of Education and Peking University to develop a plan to educate at least 5,000 students and 500 teachers within the next five years. At the pre-university level, 40 high schools have teamed up with SenseTime for the AI high school program, and there are reports of elementary schools introducing basic AI curricula to their students as well.

AI ELSI

For Canadians, China's engagement with AI from the ELSI perspective is a major concern. As a liberal democracy, the use of AI for social surveillance and militarization – especially with the explicit emphasis on civil-military technology transfer in China's official strategies – raises major

[20] Lee, Amanda. 2018. Beijing offers 1 million yuan in cash incentives, long-term visas and 'green card' to attract global talent. South China Morning Post. https://www.scmp.com/tech/china-tech/article/2138630/beijing-offers-1m-yuan-cash-incentives-long-term-visas-and-green.

[21] Cyranoski, David. 2018. China enters the battle for AI talent. Nature. https://www.nature.com/articles/d41586-018-00604-6.

[22] Tencent op. cit.

[23] See the summary of the AI Innovation Action Plan for Colleges and Universities: https://medium.com/syncedreview/china-puts-education-focus-on-ai-plans-50-ai-research-centres-by-2020-5589c35ba701.

concerns. It also seemingly justifies the call to securitize Canada's digital assets with regards to China, as the United States has done by limiting access to U.S. research institutes and funding. This is an especially difficult topic of discussion, as it is layered with cultural and political differences between the two countries, amplifying the sense of anxiety and fear that has characterized the world's response to China's rapid rise to power. In this light, it is important to consider the different ways in which China approaches AI ELSI and identify areas where constructive engagements could be pursued.

First, it is true that China uses AI for social surveillance and military applications. It regards such applications of AI as directly relating to the maintenance of stability ("social harmony") and therefore is unlikely to yield to any compromise in this area. The government plays a clear role as a client in acquiring key AI technologies for social surveillance. Since 2015, the Chinese government has been rolling out the social credit system, which rates citizens based on their behaviours within society. Based on their rating, citizens could be either punished (e.g. banned from public transit, denied loans) or rewarded (e.g. discounts on public bicycle rental fees, access to government loans). The program is expected to be rolled out nationally by 2020, and international experts have expressed concerns about the added surveillance capacity of the Chinese government and the potential implications for democratic institutions abroad as well.[24] In this light, it is difficult to imagine China being receptive to international AI ethics guidelines or standards that emphasize democratic values, such as the Montreal Declaration[25] or the European Commission's Ethics Guidelines for Trustworthy AI.[26]

It is also important to consider other approaches to AI ELSI in China. Indeed, the Party is using AI in a way that does not match the values of liberal democratic societies like Canada's – such as developing tools of social surveillance and allowing companies to access private data and test out AI products and services in a way that may jeopardize citizens' safety from the Canadian perspective. However, this does not mean that China is unaware of or indifferent to the potential dangers of AI, or that the Chinese people are passively accepting these illiberal uses by the Party. Arguably, the opposite is true.

[24] Hoffman, Samantha. 2018. Social credit: Technology-enhanced authoritarian control with global consequences. Canberra: Australian Strategic Policy Institute.

[25] Link to Montréal Declaration: https://www.montrealdeclaration-responsibleai.com/.

[26] Link to Guidelines for Trustworthy AI: https://ec.europa.eu/futurium/en/ai-alliance-consultation.

At the public level, there are concerns about AI ELSI. Chinese citizens are aware of the potential dangers of AI, as shown in a March 2018 poll that indicated 76.3 percent of Chinese people saw AI as a threat to their privacy.[27] Chinese consumers have been pushing back against misuse of private data by raising their concerns online, as well as initiating lawsuits. For the Chinese government, which greatly cares about its legitimacy, this is not an issue that it will merely disregard to meet the economic targets of the Next Generation AI Development Plan; in fact, it regards itself as the key player in protecting the collective well-being of its citizens. For instance, the Ministry of Industry and Information Technology issued warnings to Baidu, news platform Toutiao, and online payment platform Alipay for their misuse of personal data and ordered the companies to address the complaints from concerned citizens.[28]

More broadly, the Next Generation AI Development Plan clearly states the need to develop a "security assessment framework" and adjust the educational system for its citizens in preparation for a more AI-dependent economy and society. The Standardization Administration of China's AI Standardization White Paper (March 2018) points to the government's commitment to addressing ELSI as stated in the Plan. The white paper highlights the importance of safety, ethics, and privacy issues. It also demonstrates an awareness that a standardization process that takes these issues into consideration is important in order for Chinese products to remain competitive in the global market, with an eye to setting the global standards on this technology.[29] In June 2019, the Ministry of Science and Technology issued Governance Principles for the New Generation Artificial Intelligence as well (see Table 2).

Table 2: Governance principles for the new generation artificial intelligence: Developing responsible artificial intelligence[30]

- Harmony and human-friendly
- Fairness and justice
- Inclusion and sharing
- Respect for privacy
- Safety and controllability
- Shared responsibility
- Open and collaboration
- Agile governance

[27] Hersey, Frank. 2018. Almost 80% of Chinese concerned about AI threat to privacy. Technode. https://technode.com/2018/03/02/almost-80-chinese-concerned-ai-threat-privacy-32-already-feel-threat-work/.

[28] Dehua, Chi. 2018. Baidu, Alipay and Toutiao.com ordered to tackle suspected privacy violations. GBTimes. https://gbtimes.com/baidu-alipay-and-toutiaocom-ordered-to-tackle-suspected-privacy-violations.

[29] See the translation of the White Paper: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-excerpts-chinas-white-paper-artificial-intelligence-standardization/.

[30] See the translation of the principles: http://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html

Conclusion

The main takeaway of this section is that China regards AI as a key technology for its national competitiveness, and it has been investing substantial resources to achieve its goal of becoming *the* global leader in AI – in both research and application – by 2030. Regardless of the quality of China's AI innovation, the government's announcement of the Next Generation AI Development Plan has made quite a splash among the relevant stakeholders (private sector, academia, local governments) and also outside China. The perceived importance of AI for future national competitiveness, combined with the scale and speed of China's AI agenda, has generated concern and anxiety from the rest of the world, especially regarding China's seemingly illiberal application of the technology.

It is necessary to understand that China's AI ambitions are pursued in a comprehensive manner, one that does not really operate with the same clear demarcation between public and private that Canadians are familiar with. As stated in the Next Generation AI Development Plan, the development of AI in China allows technology transfer between civil and military uses. While it would be imprudent to entirely securitize Canada's research assets or businesses against China and cease engagement, it is necessary to critically approach Chinese investment in Canadian AI research institutes or commercial deals. As such, China's whole-of-the-nation approach must be taken into consideration and addressed commensurately in dealing with it on AI, whether it is on research collaboration or business engagement.

At the same time, the Government of Canada has an opportunity to positively influence China's use of this technology through smart engagement. There are no widely embraced international standards on the R&D and use of AI, and China has the ambition of being a norm-setter in this area. The relationship between China and the United States continues to become more frigid, and the European Union has become increasingly vocal in integrating explicitly liberal values into its digital policies. In this context, there is an opportunity for Canada to play the role of a middle power in finding pragmatic areas of collaboration with China, such as in the development of ethics guidelines on data and AI technologies, eschewing more sensitive topics such as social surveillance or military applications, and addressing non-political yet critical aspects of AI regulation focused on safety. China's recent softening stance on AI development and its call for international collaboration further suggests that there is such an opportunity for Canada.[31]

Overall, Canada must invest more resources in developing and implementing necessary measures to not only manage its domestic technology agenda, but also to manage its relations in the arena of foreign policy where technology and China are increasingly becoming more and more important.

[31] Knight, Will. 2018. China's leaders are softening their stance on AI. MIT Technology Review. https://www.technologyreview.com/s/612141/chinas-leaders-are-calling-for-international-collaboration-on-ai/.

JAPAN

Government Policies

AI is embedded within Japan's overall growth agenda, characterized by the vision of Society 5.0. Prime Minister Abe came to power in 2012 with an aggressive economic policy called Abenomics, characterized by measures designed to tackle economic stagnation, which had been identified as a long-term problem stemming from fundamental issues such as decreasing productivity and an aging population. Prime Minister Abe won by pledging to address these problems.

In this context, the Japanese government coined the concept of Society 5.0. This socio-economic vision aims to use emerging technologies – most notably, AI – to create "a human-centred society that balances economic advancement with the solution of social problems by a system that highly integrates cyberspace and physical space."[32] The concept was first introduced in 2016 through the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) in its fifth Science and Technology Basic Plan, Japan's five-year science and technology strategy.[33] Society 5.0 was further integrated into Japan's greater economic development strategy by the Council on Investments for the Future in June 2017, when it announced the new Growth Strategy that explicitly articulated the goal of realizing the visions of Society 5.0 and selected health care, mobility, distribution, smart cities, and financial technology as priority areas. Further, the Growth Strategy introduced the following objectives:

- Investment of government resources in strategic fields, leveraging Japan's strength;
- Development of data platforms for inter-connected utilization of data across different fields and provision of public data for private sector needs;
- Support for individual-focused adult education and strengthening of IT skills of citizens;
- Introduction of a "regulatory sandbox system"[34] to try out new ideas; and
- Linkage of the aforementioned measures to a wide range of regions, companies, and people.[35]

As part of realizing the visions of Society 5.0, the Cabinet created the Strategic Council for AI Technology to specifically address AI R&D and implementation. In this, the Strategic Council serves as a "control tower" for three main ministries relevant to AI: MEXT, the Ministry of Internal Affairs and Communications (MIC), and the Ministry of Economy, Trade, and Industry (METI).

---

[32] Society 5.0. Cabinet Office. https://www8.cao.go.jp/cstp/english/society5_0/index.html.

[33] Ministry of Education, Culture, Sports, Science and Technology – Japan. Japan's 5th Science and Technology Basic Plan. http://www.mext.go.jp/en/policy/science_technology/lawandplan/title01/detail01/1375311.htm.

[34] Regulatory sandboxes refer to testing grounds for new business models that are not protected by current regulation or supervised by regulatory institutions.

[35] 2017. Council on Investments for the Future. Prime Minister of Japan and His Cabinet. https://japan.kantei.go.jp/97_abe/actions/201705/30article6.html.

The Strategic Council co-ordinates policies within these ministries and sometimes others, such as Ministry of Health or Ministry of Agriculture, when relevant. These ministries have also been responsible for national institutes that conduct relevant research on AI, such as the National Institute of Information and Communications Technology (NICT), the National Institute of Physical and Chemical Research (RIKEN), or the National Institute of Advanced Industrial Science and Technology (AIST), which further gives them the ability to facilitate the government-industry-academia collaboration on AI.

The Strategic Council developed the Artificial Intelligence Technology Strategy in 2017, which outlines Japan's AI R&D and industrialization road map, with the objective of developing an AI industrial ecosystem by 2030. The Strategy has assigned three different research sectors for AI application (health care, productivity, and mobility) for each ministry. It has also assigned R&D agencies and private sector companies to pursue research objectives (see Table 3):

Table 3: Structure of the Strategic Council

	MIC	MEXT	METI
R&D Agencies	• National Institute of Information and Communications Technology (NICT)	• National Institute of Physical and Chemical Research (RIKEN) • Japan Science and Technology Agency (JST)	• National Institute of Advanced Industrial Science and Technology (AIST) • New Energy and Industrial Technology Development Organization (NEDO)
AI Research Area	• Natural language processing, speech translation, and brain information communication	• Basic research and infrastructure technology	• Application in industrial sectors
Research Sectors	• Health Care	• Productivity • Mobility • Health Care	• Productivity • Mobility • Health Care

The Japanese government has thus developed a framework that strategically facilitates collaboration between academia, industry, and government that leverages Japan's strengths and resources, with the ultimate objective of delivering the vision of Society 5.0.

AI R&D and Talent

Japan's self-identified weakness in AI is the lack of talent. The White Paper on International Economy and Trade (2017) noted that there is a talent shortage of approximately 50,000 in the technology field, including AI.[36] Professor Mitsuru Ishizuka of the Cognitive Innovation Center (CIC) notes that Japanese researchers are lagging behind in the field of deep neural networks, stating that Japan is a follower when it comes to this technology.

[36] Ministry of Economy, Trade and Industry – Japan. 2017. White Paper on International Economy and Trade. http://www.meti.go.jp/english/report/data/wp2017/wp2017.html.

Major corporations lead AI R&D and talent recruitment in Japan, as they have the financial resources to attract AI talent. According to a Canadian official in Tokyo, while Japanese companies have been reluctant to look outside Japan thus far, the lack of talent in AI has made them more willing to collaborate with international partners. The Perspectives on Artificial Intelligence/Robotics and Work/Employment report by Artificial Intelligence Research for Human (AIR) notes that corporations have started hiring foreign talent to address this gap.[37] The official, however, notes that most Japanese companies are looking for access to foreign AI talent through the purchase of startups abroad or on a consulting basis, as opposed to hiring and bringing them to Japan. The Japanese private sector's interest in accessing Canadian AI talent is reflected in Fujitsu's recent opening of its global AI headquarters in Vancouver, BC, in November 2018.

The government's commitment to promoting AI has had a positive impact on AI education, spurring collaboration between private and public sector entities. MEXT and METI have together created AI-related university programs, convening a national consultative body with post-secondary institutions, the Japan Business Federation, and industry stakeholders to address talent shortages and weaknesses in research.[38] Further, government ministries are collaborating closely with national research institutes and the private sector to advance research and application in the fields that they have been assigned. The Japan Deep Learning Association (JDLA) has launched an AI certificate program, in which engineers and managers go through short-term training on AI fundamentals. Lastly, Japan Science and Technology Agency also offers JPY150M to JPY500M (C$1.8M to C$6M) worth of funding in the area of AI and big data, and PRESTO for Young Researchers provides JPY30M to JPY40M (C$360,000 to C$480,000) for each three-year project for young, rising AI researchers.

AI ELSI

Society 5.0 regards economic growth and solutions for social problems as inseparable. Thus, in addition to R&D and economic growth, Society 5.0 includes Japan's approach to ELSI in relation to AI development. The Cabinet's commitment to ELSI is most notable in the establishment of the Advisory Board on Artificial Intelligence and Human Society.[39] The Advisory Board consists of 12 members from different academic backgrounds (humanities, social sciences, engineering, law, etc.) and is designed to advise the Cabinet on ELSI matters related to AI from an interdisciplinary perspective. The Board's Report on Artificial Intelligence and Human Society (March 2017)

[37] Ministry of Education, Culture, Sports, Science and Technology – Japan. Acceptable Intelligence with Responsibility, Perspectives on Artificial Intelligence/Robotics and Work/Employment. http://www.mext.go.jp/en/policy/science_technology/lawandplan/title01/detail01/1375311.htm.

[38] Harris, Paul. 2017. Research brief: Developments in artificial intelligence (AI) in Japan and implications for Australia. Department of Education and Training. https://internationaleducation.gov.au/International-network/japan/PolicyUpdates-Japan/Documents/AI%20in%20Japan%20research%20brief_07-2017.pdf.

[39] Ema, Arisa. 2017. EADv2 Regional Reports on A/IS Ethics: Japan. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_regional_report.pdf.

addresses concrete, imminent issues in mobility, manufacturing, personal services, and communication from ethical, legal, economic, educational, social, and R&D perspectives. Further, MIC organized the Conference toward AI Network Society in 2017, which convened approximately 40 members from academia, industry, and civil society. A subcommittee at the conference drafted AI R&D principles, an ethics guideline for researchers. It is clear that the Japanese government is attentive to groups concerned about AI ELSI and provides space for active discussions.

Table 4: AI R&D principles[40]

- Principle of collaboration: Developers should pay attention to the interconnectivity and interoperability of AI systems (principles mainly concerning mitigation of risks associated with AI systems).
- Principle of transparency: Developers should pay attention to the verifiability of inputs/outputs of AI systems and the explainability of their judgments.
- Principle of controllability: Developers should pay attention to the controllability of AI systems.
- Principle of safety: Developers should take it into consideration that AI systems will not harm the life, body, or property of users or third parties through actuators or other devices.
- Principle of security: Developers should pay attention to the security of AI systems.
- Principle of privacy: Developers should take it into consideration that AI systems will not infringe on the privacy of users or third parties.
- Principle of ethics: Developers should respect human dignity and individual autonomy in R&D of AI systems (principles mainly concerning improvements in acceptance by users, etc.).
- Principle of user assistance: Developers should take it into consideration that AI systems will support users and make it possible to give them opportunities for choice in appropriate manners.
- Principle of accountability: Developers should make efforts to fulfil their accountability to stakeholders, including AI systems' users.

Lively discussions on AI ELSI have taken place in academia as well. The Japanese Society of Artificial Intelligence (JSAI) established its Ethics Committee in 2014. The Ethics Committee started drafting a Code of Ethics in early 2016 and published the Japanese Society for Artificial Intelligence Ethical Guidelines in May 2017. The Guidelines, directed at AI researchers, stress the importance of being aware of their social responsibilities and maintaining effective communications with society (see Table 5).

Table 5: JSAI Ethical Guidelines[41]

- Contribution to humanity
- Abidance of laws and regulations
- Respect for the privacy of others
- Principle of fairness
- Principle of security
- Acting with integrity
- Accountability and social responsibility
- Communication with society and self-development
- Abidance of ethics guidelines by AI

[40] Ministry of Education, Culture, Sports, Science and Technology – Japan. Acceptable Intelligence with Responsibility, Perspectives on Artificial Intelligence/Robotics and Work/Employment. http://www.mext.go.jp/en/policy/science_technology/lawandplan/title01/detail01/1375311.htm

[41] Ibid.

There are also other groups, such as the JDLA and AIR. As such, there are forums in Japan for interdisciplinary scholars to raise AI ELSI matters to policy-makers as large-scale domestic policies are formulated. AI ELSI continues to raise concerns around the world. In this context, Japan, a country with a great AI ambition, the technological capability to deliver it, and similar political values as Canadians, stands out as a promising partner in the area of international governance in matters of AI ELSI.

Japan has started mobilizing its resources to play an active role in shaping the international governance of AI. During his policy speech delivered to the National Diet on January 28, 2019, Prime Minister Abe announced that Japan "will take the lead in establishing human-centred ethical principles for AI."[42] His administration has thus far followed through with this pledge. Japan has agreed to collaborate more closely with Canada and the European Union on AI during Prime Minister Abe's trips in April and May, with an emphasis on promoting "human-centric" uses of the technology.

Conclusion

Strong, centralized leadership by the government underscores Japan's AI strategy. The policies that the Japanese government has rolled out thus far make it very clear that it regards increasing its R&D capacities and integrating AI into its economy – particularly the manufacturing sector – a key priority within the national economic agenda. The formation of the Strategic Council – the control tower – underscores the importance of the government support. The Council creates the space for key ministries – MEXT, METI, and MIC – to co-ordinate their AI policies, while also providing a forum for policy-makers, industry, and research to collaborate in creating the foundation for a new economy driven by emerging technologies such as AI.

The emphasis on AI within Japan's greater economic policy has spilled over to the private sector in the area of talent. The government and the industry are keenly aware of Japan's lag in basic AI research, and they are looking at Canada to fill the gap. Japanese companies, known for their conservative attitude toward foreign workers, are now realizing that they need to work with those abroad in order to complement their weakness in AI research. Japan's capital and manufacturing capabilities could be effectively coupled with Canada's strength in AI basic research.

Finally, the Japanese government's proactive movement in AI ELSI again underscores the degree of importance that it has invested in AI. The fact that the government has invested in internal bodies such as the Advisory Board on Artificial Intelligence and Human Society highlights the Japanese government's understanding that AI cannot be confined to the realm of science and technology policy, but that of the greater national development policy. Further, Japanese AI ELSI bodies have been active in producing early and widely distributed AI ethics guidelines, with an eye to the international governance of the technology.

[42] Cabinet Public Relations Office. 2019. Policy Speech by Prime Minister Shinzo Abe to the 198th Session of the Diet. https://japan.kantei.go.jp/98_abe/statement/201801/_00003.html.

Overall, the review of Japan's AI strategy demonstrates that Japan, like China, considers AI and its integration into the economy as a priority within its national economic agenda, and that it has been making use of its powerful government bureaucracy to spur the collaboration between policy-makers, industry, and academia. Japan's strategy is comprehensive and co-ordinated, viewing AI as a key marker of global competitiveness for Japan in the world.

In this context, Japan stands as an attractive partner for collaboration in the space of AI. Canada's strength in basic research complements Japan's strength in manufacturing and application, as well as abundance of capital. Further, the alignment of values between Japan and Canada suggests that the two could also collaborate in the promotion of global AI governance. Being mindful of Japan's substantial commitment to AI, the Government of Canada should look to Japan for a like-minded partner.

SOUTH KOREA

Government Policies

In 2017, the Ministry of Science, ICT, and Future Planning (now the Ministry of Science and ICT) launched the Mid-to-Long-Term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution. The policy document addresses R&D strategies and potential ELSI of greater use of AI and introduces a road map for achieving the following goals:

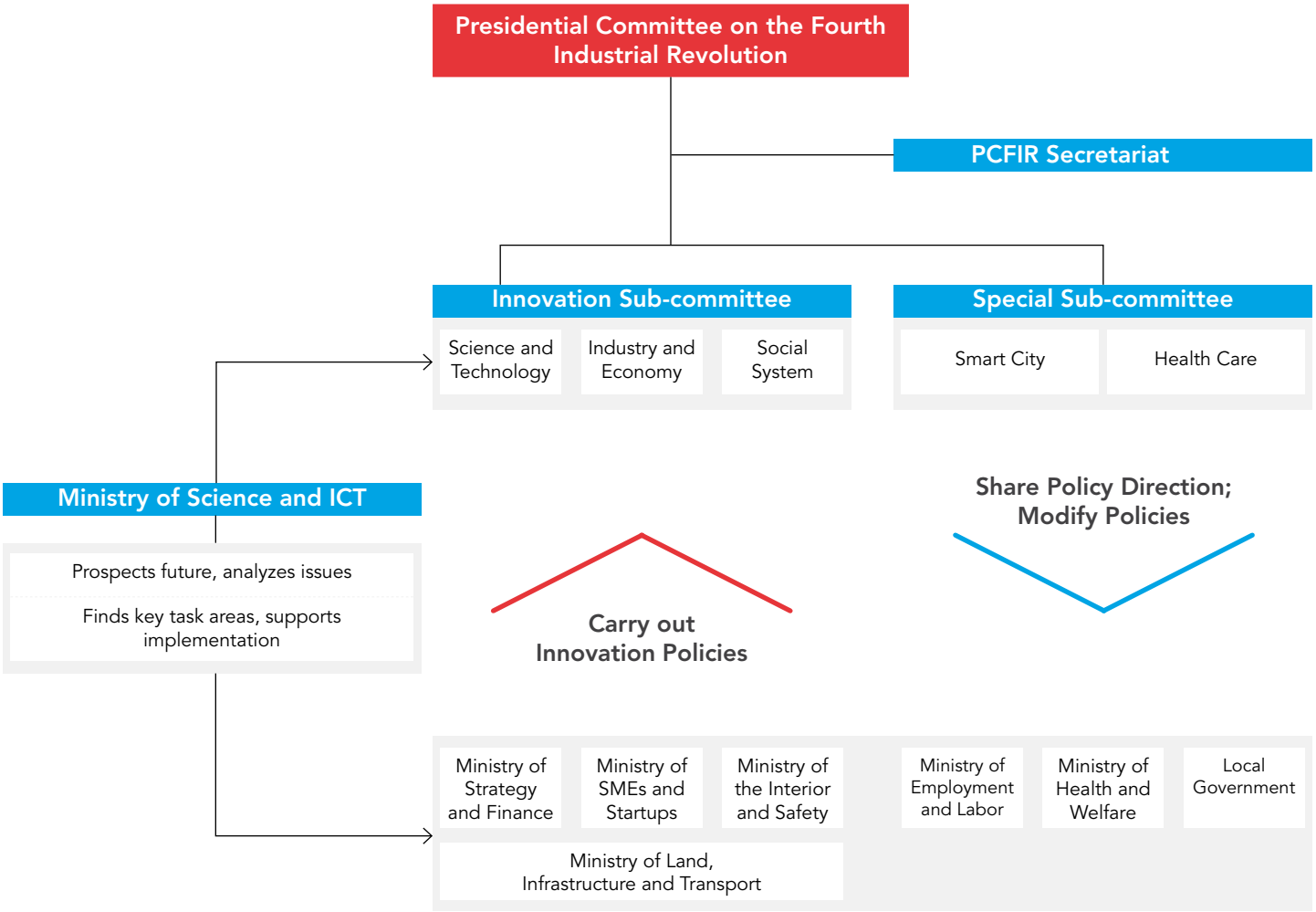
- Build a world-class technological foundation;
- Promote intelligent industry[43]; and
- Revise existing social policies and regulations.

The document provides a framework for collaboration between technology, industry, and civil society with a 30-year time frame. The Ministry of Science, ICT, and Future Planning was re-structured to the Ministry of Science and ICT with the change of power in May 2017, and the Mid-to-Long-Term Master Plan has become outdated. However, the Plan reflects the South Korean government's basic approach to AI, which is re-articulated in more detail through the Presidential Committee on the Fourth Industrial Revolution (PCFIR).

Newly elected President Moon Jae-in established the PCFIR in November 2018. Operating under the motto "led by the private sector, supported by the government," the PCFIR features a roster of private sector leaders and academics alongside five ministers from relevant departments and the science advisor to the President, totalling 25 members (see Table 6). Approximately 30 staff members from the Ministry of Science and ICT support the day-to-day operations of the PCFIR.

[43] When South Korean documents refer to "intelligence" or "intelligent," it often refers to AI in particular.

Table 6: Composition of the PCFIR



Source: PCFIR Website

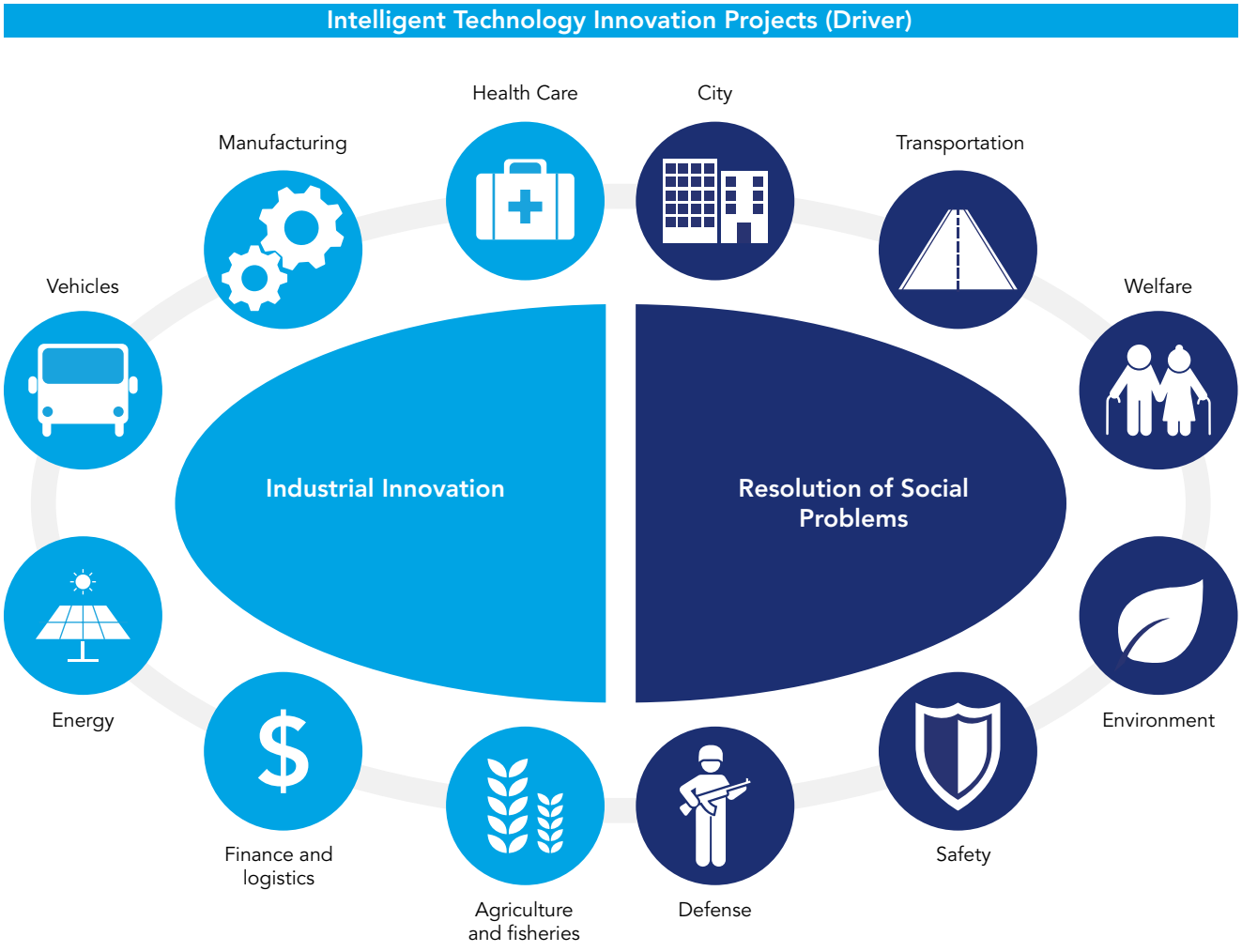
Similar to Japan's Strategic Council for AI Technology, the PCFIR serves as a control tower for co-ordinating the government's policies on AI and other emerging technologies. More specifically, the mandate of the PCFIR is the following:

- Co-ordination of policy measures submitted by various ministries and the committee members;
- Organization of public campaigns related to the fourth industrial revolution and encouraging public participation;
- Preparing the groundwork for regulatory and institutional reforms in support of public-private partnerships; and
- Fostering ecosystems for emerging industries.

The PCFIR's key document is the People-Centered Response Plan for the Fourth Industrial Revolution to Promote Innovative Growth, or I-Korea 4.0 Strategy. The Strategy outlines a comprehensive national strategy that seeks to support R&D and deploy the technologies of the fourth industrial revolution. I-Korea 4.0 parallels Japan's Society 5.0 in that both outline a comprehensive vision in which the technologies of the fourth industrial revolution are developed and deployed to support economic growth and solve social problems.

More specifically, I-Korea 4.0 outlines the plan to promote intelligent technology innovation projects in 12 different sectors (see Table 7); secure growth engine technologies; create industrial infrastructure and ecosystems; and make preparations for future social change entailing the use of disruptive technologies. In I-Korea 4.0, the PCFIR aggregates and co-ordinates proposed policies from relevant ministries, with the Ministry of Science and ICT taking the lead. The time frame for the Strategy is five years, from 2018 to 2022, and the Strategy lists concrete objectives for all 12 sectors.

Table 7: The 12 sectors in I-Korea 4.0



Source: I-Korea 4.0 booklet

In June 2019, President Moon announced a manufacturing renaissance strategy, which will inject KRW8.4T (US$7.1B) in key industries such as non-memory chips, future mobility, and biohealth technologies, as well as upgrade South Korea's manufacturing sector through integration of AI. The manufacturing renaissance strategy echoes Made in China 2025 in many ways, articulating South Korea's ambition to not only maintain, but also enhance, its manufacturing competitiveness, striving to become one of the top four global exporters, as well as increasing the value-added ratio of the manufacturing sector from 25 percent to 30 percent by 2030.

The strategy is not a deviation from I-Korea 4.0; rather, it builds on I-Korea 4.0 as a blueprint and fleshes out the details specifically within the manufacturing sector. In addition to the construction of 30,000 smart factories outlined in I-Korea 4.0, the government will also support the development of 2,000 "AI factories" by 2030. The process will also include the establishment of a data centre for smart factories and revision of existing laws to enhance the business environment for the manufacturing sector. Further, the Ministry of Industry will take a lead in drafting an AI national strategy on manufacturing within 2019.

AI R&D and Talent

South Korean policy-makers and AI practitioners generally believe that their AI R&D sector is underdeveloped. A common refrain in policy documents and conversations held with AI experts in Korea is the fear of falling behind or the need to catch up to the rest of the world. The lack of AI talent is a major policy concern for South Korean policy-makers and the industry, which has prompted the introduction of the national AI R&D Strategy.

The PCFIR launched the AI R&D Strategy in May 2018, investing KRW2.2T (C$2.66B) until 2022 to acquire world-class AI technology, develop AI talent, and become one of the top four AI nations. More specifically, the AI R&D Strategy includes the following initiatives:

- Government-funded AI projects in public sectors (defence, health care, public safety) modelled after DARPA Grand Challenge competitions in the United States;

- Creation of six additional AI graduate programs and five AI R&D institutes for training 5,000 new AI specialists by 2022;
- Promotion of AI application in promising industries (e.g. pharmaceutical research); and
- Creation of an AI hub to provide data relevant to local entrepreneurs and industries.

Adding to the challenge is the disparity in resources between the conglomerates and small and medium-sized enterprises (SMEs). Startups and SMEs currently compete against multinational conglomerates (e.g. Samsung, LG) that are not only acquiring talent from within, but also from the rest of the world.

Table 8: AI research centres of South Korean Multinational Corporations

Corporate AI Research Labs	Year of Establishment
Naver Labs AI	2015
SK Telecom AI	2016
Samsung AI Center	2016
LG AI Lab	2017
SK T-Brain	2017
Kakao Brain	2017
Naver Clova	2017
Samsung SDS AI	2017
Hyundai Motors AI	2017

South Korean conglomerates have started looking abroad to access top AI talent around the world. Naver (often referred to as South Korea's Google) purchased XRCE in France to gain access to 80 top European AI experts. The conglomerates have also been active in Canada. Samsung opened its second North American AI Centre in Toronto in May 2018, following the launch of its AI lab in Montreal in partnership with the Université de Montréal's MILA (one of the three AI clusters) in September 2017. LG also established a partnership with the University of Toronto – home to the Vector Institute, another AI cluster – to set up an AI research lab.

AI ELSI

ELSI is integrated into South Korea's blueprint for its fourth industrial revolution policies. The PCFIR's objective of using cutting-edge technologies to boost economic growth and solve social problems makes it very clear that ELSI is a central concern for South Korea's lawmakers. I-Korea 4.0 specifically points to policies of integrating AI into health care, welfare, the environment, and public safety. One of the PCFIR's three subcommittees is the Social Institutions Subcommittee, which is tasked with innovation of employment and welfare policies, educational reforms to foster creativity, legal and policy reforms to solve real social problems, and international and community relations.

One of the four pillars of I-Korea 4.0 is to make preparations for future society. The PCFIR lists the following policy measures that fall within this pillar:

- Broadening of scope for employment insurance and strengthening of the social safety net for job retraining;
- Reflecting the changing nature of the work environment to ensure that workers in new industries are eligible for workers' compensation;
- Creating an ethics guideline for AI;
- Creating measures to ensure that those negatively impacted by AI have a legal basis for proper compensation; and
- Continuing to hold multi-stakeholder discussions through the PCFIR to create "social consensus" on the government's response to the fourth industrial revolution on labour, education, the social safety net, and ethics.

One of the signature events of the PCFIR is the Regulatory and Institutional Reform Hackathon, which has been held 13 times since the launch of the PCFIR. Modelled after sprint-like, intense work sessions among software engineers in the tech sector, the Hackathon has brought together approximately 30 stakeholders from government, civil society, and the private sector for two-day workshops. The participants discuss multi-stakeholder issues, such as data and privacy rights, the data cloud, or the drone industry in an open manner, curated by professional facilitators. With an emphasis on consensus making, the participants create policy recommendations to the PCFIR, relevant ministries, and the National Assembly. The progress on their recommendations is reported to all participants quarterly. While not all of the discussions touch on AI, these hackathons provide a model for bringing together relevant stakeholders and influencing policy-making in a nimble, democratic manner.

Table 9: Regulatory and Institutional Reform Hackathons

Date	Themes
December 2017	• Right to Self-determination in Financial Information • Improving Location Information Protection Law • Improving Advanced Medical Equipment Regulation
January 2018	• Improving Certification Systems • Harmonizing Use and Protection of Personal Information
April 2018	• Harmonizing Data Use and Protection of Personal Information • Promoting Drone Industry • Promoting Use of Cloud Computing in Public Sectors
September 2018	• Innovation in Transportation Service through ICT • Legalizing Shared Rentals for Korean Citizens in Urban Areas • Improving Regulation of Combined Products in Health Care
March 2019	• Eliminating Regulatory Grey Zones Created by Increase of Personalized Transport • Stimulating Food Industry by Improving Labelling of Nutritional Functionality

On the ethics front, the Ministry of ICT, in collaboration with the National Information Society Agency, released the Ethics Guideline for Intelligent Society in June 2018, as indicated by I-Korea 4.0. The Guideline builds on the principles of Seoul PACT (Publicness, Accountability, Controllability, and Transparency) (see Table 10).

Table 10: Seoul PACT

Publicness: Intelligent information technology should benefit as many people as possible, and the economic benefits produced by the technology should be shared widely for the prosperity of humanity.
Accountability: Clarify the distribution of liabilities entailed to intelligent information technology and related services, and ensure that the distribution of information on safety and protection of user rights and other related social duties is undertaken.
Controllability: Pre-emptively prepare responses to malfunctions of intelligent information technology and services, and ensure that user rights for choices be guaranteed as much as possible.
Transparency: R&D and design should strive to reflect opinions of users, consumers, and citizens; disclose potential dangers while in use; and ensure that personal information is processed appropriately.

Conclusion

South Korea's rapid economic development following the ravages of the Korean War has been referred to as the Miracle on the Han River and became a benchmark for developing nations around the world. The miracle was made possible through economic policies in which the government and private sector entities collaborated closely with each other, taking a whole-of-the-nation approach to economic and social development. The catch-up mindset of the twentieth century is reflected in South Korea's approach to AI and, more broadly, the fourth industrial revolution. Like China and Japan, South Korea regards AI as a key driver of the next generation economy and therefore a national project, closely tied to its global competitiveness. As such, the government has taken measures to ensure that South Korea does not fall behind the rest of the world, clearly articulated in its goal of becoming one of the top four AI nations by 2022. Canada's C$125M investment in its national AI strategy pales in comparison to South Korea's C$2.66B investment in its R&D strategy, and even more so considering other policy measures taken within the PCFIR's greater agenda.

South Korea and Canada have strengths that complement each other. Experts have noted Canada's strength in basic AI research, a product of a long-term investment in its scientists; meanwhile, South Korea excels in implementation and has hardware capacities through its strong manufacturing sector. South Korea's emphasis on upgrading its manufacturing sector through integration of AI is consistent with the objectives of China and Japan, and also highlights the same opportunity for Canada: Canadian AI talent and the East Asian partner's manufacturing capacity. The rush of South Korean conglomerates to Canada is a testament to this potential, and Canada should explore ways of harnessing this opportunity even further.

Finally, the examination of South Korea's approach to AI underscores the potential for collaboration between Canada and South Korea on the issues of AI ethics and broader politics of AI governance. The South Korean government is aware of – and has taken policy measures through the PCFIR to act upon – the potential ELSI of AI. When it comes to values, Canada and South Korea have more in common than not. Further, in light of the growing chasm between China and the United States, Canada and South Korea have the potential to collaborate closely on advocating for the ethical use of AI at the international level.

ANALYSIS & POLICY RECOMMENDATIONS

Analysis

This review of AI policies in East Asia suggests that all three governments regard AI as a transformative technology that is critical to their economic and social development. They regard the state as a key actor in delivering this AI vision, and thus have created comprehensive strategies that stretch beyond R&D and talent development to consider AI's integration into their economy and measures to address associated ELSI.

Accordingly, they have created AI-specific bodies, such as China's AI Plan Promotion Office under the Ministry of Science and Technology; Japan's Strategic Council for AI Technology; and South Korea's Presidential Committee on the Fourth Industrial Revolution, to co-ordinate AI-related endeavours of different ministries and facilitate a coherent, long-term, national-scale plan of action.

These AI strategies create a framework in which a set of goals on AI development are outlined, and the private sector, with the policy support of the state, delivers them. More specifically, these strategies create a space where policy-makers, private enterprises, and academia can collaborate closely. South Korea's strong bureaucracy and co-ordinated policy-making allow the state to play the role of facilitator among different stakeholders. Given that AI requires not only advances in R&D and capital, but also adjustment of existing policies and regulations at the government level, this comprehensive, national approach to AI will facilitate a favourable environment for the East Asian states to move forward in the AI space.

China, Japan, and South Korea believe that they are lagging behind in R&D and talent development, and therefore have prioritized these perceived deficiencies. They have enacted policies to address the shortage of AI talent, by both creating education programs and attracting foreign talent. Their private sectors are also aggressively recruiting and creating research centres abroad to access foreign AI talent. In this, they have shown considerable interest in gaining access to Canada's AI ecosystem and talent.

Further, these government bodies and plans point to their awareness of ELSI resulting from increased deployment of AI. China's Next Generation AI Development Plan and South Korea's I-Korea 4.0 address the need to both adjust social and legal institutions in place and adequately prepare their citizens in the long run. Similarly, the Advisory Board in Japan has made recommendations on ELSI to the Cabinet through its Report on Artificial Intelligence and Human Society. All three East Asian governments display an awareness of the potential dangers of AI and a willingness to engage with the ELSI issues in both development and deployment with a long-term perspective.

AI development in each of the three states is considered a national project. The strategies of each country conform to the developmental state theory, which explains the past success of East Asian states in achieving rapid economic growth by appealing to the nationalist rhetoric that they

are falling behind the rest of the world and thereby justifying significant state involvement in economic planning.[44] With the rise of AI geopolitics, therefore, it is important to understand that AI development is regarded as a national project, and Canada's engagement in this space – whether the objectives are to promote trade or address national security concerns – requires a more comprehensive approach, especially with a technology like AI, which blurs previously set sectoral boundaries.

Finally, on values: it is important to understand that substantive discussion on AI ethics will have to go beyond broad statements featured in respective ethics guidelines. Table 11 demonstrates that ethics guidelines from the three East Asian states and Canada do not deviate much from each other. These guidelines share the idea that AI must be used for some sort of greater good, and that it must be used in a way that is safe, transparent, or responsible. From a broader point of view, the difference lies in the specific social, economic, political, and cultural context in which AI systems are deployed. This difference is an implicit premise of this report.

[44] See Leftwich, Adrian. 1995. Bringing politics back in: Towards a model of developmental state. The Journal of Development Studies. 31(3).

Table 11: Comparison of ethical guidelines

Key Themes	**Canada** (Montreal Declaration)	**China** (Principles of Next Generation AI Governance)	**Japan** (JSAI Ethical Guidelines)	**South Korea** (Seoul Pact)
Privacy	• Privacy and intimacy	• Respect for privacy	• Respect for the privacy of others	
Justice	• Equity	• Fairness and justice	• Principle of fairness	
Safety	• Prudence	• Safety and controllability	• Principle of security	• Controllability • Transparency
Responsibility	• Responsibility	• Shared responsibility	• Acting with integrity • Accountability and social responsibility	• Accountability
Society and Governance	• Solidarity • Democratic participation • Diversity inclusion • Sustainable development	• Inclusion and sharing • Agile governance	• Abidance of laws and regulations • Communication with society and self-development	• Publicness
Human-centric well-being	• Well-being • Respect for autonomy	• Harmony and human-friendly	• Contribution to humanity • Abidance of ethics guidelines by AI	
Other		• Open and collaboration		

This has several important implications. Shared values underline the fact that the concern for responsible AI transcends political differences, and that there is space for constructive, transnational dialogue on AI governance with the three East Asian states, at least in theory.

However, prior experience indicates that beyond shared values articulated in documents, context, which affects their interpretation, implementation, and enforcement, matters more. In this situation, it is important for Canada to continue engaging with its partners on AI, finding common ground in the actual practice of these ethical values through communication and collaboration. It is easier and more expedient to argue for this approach in relation to Japan and South Korea, but not so much with regards to China, which has increasingly become the "other" of the Western, liberal states. But especially for this reason, it is important for states like Canada to continue engaging with China, in order to ensure that important dialogues on AI governance (and beyond) are not siloed along geopolitical fault lines; and to constantly refine and identify areas of dialogue and negotiation, focused on issues of mutual concern.

Policy Recommendations

Recommendation 1: Increase Canada's capacity to govern AI in a co-ordinated manner at different levels of government.

Canada stands to benefit from the Government of Canada's forward-looking investment in academic research in the field of AI. The C$125M investment in the Pan-Canadian Artificial Intelligence Strategy will further boost Canada's excellence in research, striving to maintain its global competitiveness. However, this evaluation of East Asian AI strategies suggests that there is a need for a more comprehensive Canadian AI strategy that stretches beyond R&D and that would harness Canada's capacity to adequately compete with better co-ordinated states in AI like China, Japan, or South Korea.

These countries have recognized that AI is a technology that is not limited within a single domain, and that engagement with it requires a cross-ministerial effort, which requires expertise to govern. Therefore, they have created bodies such as the AI Plan Promotion Office (China), Strategic Council (Japan), and the Presidential Committee on the Fourth Industrial Revolution (South Korea). In May 2019, the Government of Canada created the Advisory Council on Artificial Intelligence to "advise the Government of Canada on how best to build on Canada's AI strengths, identify opportunities to create economic growth that benefits all Canadians and ensure that AI advancements reflect Canadian values."[45] This is a new development and not much information is available, but it is definitely a step to the right direction.

Canada is different from East Asia and the developmental state approach would not be applicable in the context of Canada's system of governance, especially considering different levels of government. It is challenging for the Canadian government to engage with better co-ordinated Asian states, because it does not have the same capacity to centralize its AI agenda as they do. This is a challenge that has no single answer, but the Government of Canada must lead and hold dialogues with lower-level governments to ensure that AI is deployed across the country and to formulate AI policies that balance the need for co-ordination while maintaining the respect for existing relations between different levels of government.

Increasing the state capacity to govern AI starts by creating a space and investing in talent that can facilitate agile discussions and policy implementations in co-ordination with different levels of government and non-governmental stakeholders. While it is difficult to imagine a centralized AI agency with the same capacities as the ones in East Asia, it might be reasonable to create an AI hub for federal and provincial AI R&D and deployment strategies in different sectors, as well as policy guidelines and public engagement.

[45] Innovation, Science and Economic Development Canada. 2019. Government of Canada creates Advisory Council on Artificial Intelligence. News Wire Canada. https://www.newswire.ca/news-releases/government-of-canada-creates-advisory-council-on-artificial-intelligence-838598005.html.

There are great initiatives rolling out in different levels of government in addition to the Pan-Canadian Strategy, such as the federal government's Innovation Superclusters Initiative or the Alberta government's recent decision to invest C$100M to help AI sectors. Also, the federal government has worked with the Council of Chief Information Officers to set standards for use of AI and hosted National Data and Digital Consultations roundtables to address the ELSI issues. The creation of a central AI agency would allow the different levels of government to work together more effectively to maximize the potential of different AI-related programs, ensure that the diffusion and adoption of the technology across the country follow common norms and ethics, and provide the capacity for Canadian policy-makers to lead in AI governance at the international level.

Recommendation 2: Leverage Canada's AI talent more effectively as part of "Brand Canada."

Canada is home to world-class researchers strong in basic AI research. This is an asset particularly important at a time when the global AI talent shortage is the most common, pressing challenge for governments seeking to harness their AI potential. The Pan-Canadian AI Strategy recognizes this strength and seeks to maintain this competitive edge.

However, at least in East Asia, there is a need to highlight the association between these world-class researchers with the Canada brand. While many recognize the names of researchers, they do not necessarily associate them with Canada. Although major corporations from East Asia (e.g. Samsung, Baidu, Fujitsu) are rushing to establish research centres to gain access to Canada's AI talent, more could be done to attract investment to Canada. Canadian governments and AI stakeholders should work together on branding Canadian AI to further stress the association with Canada, thereby creating opportunities for post-secondary universities and local economies that host these world-class researchers.

More specifically, it is necessary for Canada's AI stakeholders – federal and provincial governments, academics, and practitioners – to identify what makes Canadian AI distinct from U.S., Chinese, or any other competitor's AI. Initial conversations have suggested that Canada's diversity and excellence in good governance are some of the qualities that should be stressed, and a policy that combines immigration – a specific policy area that reflects these Canadian values – with AI talent acquisition might be a productive way forward.

Finally, it is necessary to harness the economic benefit of Canadian AI beyond R&D and connect it to its application to different sectors of the Canadian economy. The Innovation Superclusters Initiative is a positive measure that enhances existing areas of strength in Canada's diverse economy and showcases it to the rest of the world. In branding Canadian AI, it is necessary to integrate a broader picture that goes beyond highlighting excellence in R&D.

Recommendation 3: Proactively address emerging AI issues in trade and investment.

Due to the nature of the technology, advances in AI development and deployment will create new issues in trade and investment with China, Japan, and South Korea, which are Canada's second, fifth, and seventh largest trading partners, respectively. As such, Canada's different levels of government should consider emerging AI issues in trade and investment and prepare a future course of action.

Canadian AI entrepreneurs are aware of the market opportunity in China but are apprehensive about the issues of intellectual property regulations and potential transfer of their technology for social surveillance or AI weapons. This apprehension stems from the perception of different

approaches to AI ethics, a problem that will not necessarily be limited to China. Further, cross-border data regulation, which is tied to AI development and deployment, has already been identified as an emerging trade issue. In this context, different governments in Canada – especially the federal government – must harness the capacity to respond to issues related to AI and trade in an efficient and timely manner.

Canadian AI talent and research centres continue to attract investments from major East Asian companies. AI is a permeable technology and could pose threats in different social, economic, and political sectors. This does not mean that Canada should resort to isolationism driven by fear; on the contrary, it is necessary for policy-makers to fully assess the implications of foreign investments in the space of AI and create clear guidelines for all stakeholders.

Finally, it should be noted that when it comes to AI policy, there is no silver bullet that will adequately address all issues at once. The focus should be on creating a space for continuous discussions and devising policy responses in a nimble manner, instead of drafting heavy handed regulations that cannot keep pace with the technology.

Recommendation 4: Identify Canada's place in emerging AI geopolitics and engage strategically.

China's rapid rise in the field of AI has created a sense of anxiety around the world, coupled with the existing fears toward China. Conversations about AI in the Western world increasingly stress points such as democracy and civil rights, which read as a response to China's use of AI for social surveillance. Key discussions on AI ethics and policy occur in forums such as the Organisation for Economic Co-operation and Development (OECD) or the G7, which exclude China. Consequently, a fracture seems to be emerging between China and the Western world, which may prevent constructive dialogue on AI ethics and global governance. This fracture seems to have deepened significantly and become more obvious since September 2018, when the first draft of this report was developed.

Japan and South Korea share apprehension about China's AI leapfrogging, viewing it as a threat to their economic competitiveness. Politically, they are considered to be part of the liberal countries, but there are concerns about an overly Western imposition of AI ethical discussions, stemming from distinct cultural and intellectual traditions. For instance, even though both are liberal democracies, Japan and South Korea's discussions of AI ethics avoid explicit use of liberal language, such as democracy or civil rights. Further, for states like Japan and South Korea, their proximity to and economic dependence on China will increasingly limit their full commitment to causes of the liberal countries.

In this context, Canada is in a position to play a positive role on the international stage as a mediator. It has strong AI researchers and institutions as an asset and the credibility that China, the United States, and European countries lack vis-à-vis each other. Regrettably, Canada's relations with China have deteriorated significantly since the arrest of Huawei CFO Meng Wanzhou, but with the right pivoting in the long run, Canada could be the player that brings these parties to the table to facilitate dialogue on AI ethics and global governance.

In sum, the Government of Canada must embrace AI as a space of international competition among states, an amalgam of the geopolitical and technological. In addition to continuing to harness its AI capacity, it should also actively participate in AI ethics and governance moving forward.

ACKNOWLEDGEMENTS

I am grateful to everyone who agreed to share their knowledge and thoughts during this research. I would like to particularly note the following individuals and organizations for their help and valuable discussions during the field research in 2018:

- Beijing: Professor Roger Ames (University of Hawaii), the United States Information Technology Office, and the Embassy of Canada to China in Beijing.
- Tokyo: Professor Arisa Ema (University of Tokyo), Professor Mitsuru Ishizuka (University of Tokyo), and the Canadian Embassy in Japan.
- Seoul: The Office of Presidential Committee on the Fourth Industrial Revolution, the Artificial Intelligence Industrial Association, Professor Jung-won Lee (University of Seoul), and the Canadian Embassy in Seoul.

The research, including the fieldwork, was principally funded by the Asia Pacific Foundation of Canada's Post-Graduate Research Fellowship. The completion of the project would have been impossible without the support of the Foundation. I would also like to particularly acknowledge the following individuals at the Foundation (in alphabetical order of their last names):

- Sanya Arora
- Eva Busza
- Justin Elavathil
- Scott Harrison
- Serena Ko
- Charles Labrecque
- Sebastian Murdoch-Gibson
- Michael Roberts
- Pauline Stern
- Kai Valdez-Bettcher

In addition, part of the material in this report was produced for the International Policy Ideas Challenge funded by Global Affairs Canada and Social Sciences and Humanities Research Council of Canada, Government of Canada.

ABOUT THE ASIA PACIFIC FOUNDATION OF CANADA

The Asia Pacific Foundation of Canada (APF Canada) is dedicated to strengthening ties between Canada and Asia with a focus on expanding economic relations through trade, investment, and innovation; promoting Canada's expertise in offering solutions to Asia's climate change, energy, food security, and natural resource management challenges; building Asia skills and competencies among Canadians, including young Canadians, and improving Canadians' general understanding of Asia and its growing global influence.

Visit APF Canada at http://www.asiapacific.ca.

ASIA PACIFIC FOUNDATION OF CANADA
FONDATION ASIE PACIFIQUE DU CANADA