AI Decision-making in Smart Cities, Japan's Society 5.0

Ruben E. Rodriguez Samudio

A. Introduction

Population trends reveal that citizens are increasingly moving to cities. According to the UN, in 2018, 55.3% of the world's population (4.4 billion people) lived in cities. By 2030, this number will rise to 60.4%.[1] Furthermore, the higher the income of a particular country, the higher percentage of its population that chooses to live in an urban environment.[2] Therefore, the importance of city development goes hand in hand with a country's economic and population growth. This phenomenon is not limited to large cities. A disproportionate number of fastest-growing cities are in the lower-middle-income range.[3] As this trend continues, cities will require better administration to accommodate the population increase if this trend continues.

[1] United Nations, "World Urbanization Prospects: The 2018 Revition", Department of Economic and Social Affairs, 2019, https://population.un.org/wup/publications/Files/WUP2018-Report.pdf.
[2] OECD/European Comission, *Cities in the World: A New Perspective on Urbanization* (Paris: OECD Publishing, 2020), https://doi.org/10.1787/d0efcbda-en.
[3] World Bank, *Competitive Cities for Jobs and Growth: What, Who, and How*, (Washington: The World Bank Group, 2015), https://documents1.worldbank.org/curated/en/902411467990995484/pdf/101546-REVISED-Competitive-Cities-for-Jobs-and-Growth.pdf.

DOI: https://doi.org/10.17875/gup2023-2499

Smart cities are becoming the preferred model of urban development to face the challenges of an increasingly urban population. Governments utilize Information and Communications Technologies (ICT), namely the Internet of Things (IoT), big data analytics, and cloud computing, to collect, transfer, process, and store data. *Edwards* observes that in developed countries, smart cities tend to be "retrofitted" to meet environmental, social, political, or business targets. By contrast, in developing countries, cities are "created from scratch" to enable modernization and development to address population, climate, and migratory issues.[4]
Japan introduced its approach to smart cities in its 5th Science and Technology Basic Plan[5]. Under the title Society 5.0, the Japanese government aims to create a smart society, defined as:

> "[A] *society where the various needs of society are finely differentiated and met by providing the necessary products and services in the required amounts to the people who need them when they need them, and in which all the people can receive high-quality services and live a comfortable, vigorous life that makes allowances for their various differences such as age, sex, region, or language."*

Even before the unveiling of the Society 5.0 plan, there were over 200 smart city projects in Japan.[6] In addition, Japanese industry giants are also actively developing and implementing new initiatives. For example, Panasonic is developing "sustainable smart towns" in existing cities. The first one, Fujisawa SST, was created in 2014 on one of its closed factory sites.[7] Similarly, Toyota's Woven city, currently under construction in Susono City, Shizuoka, is a 708,000 m^2 project fueled by hydrogen cells and is envisioned as a "living laboratory." Its purpose is to serve as a home to full-time residents and researchers who will be able to test and develop technologies such as autonomy, robotics, personal mobility, smart homes, and artificial intelligence in a real-world environment."[8]

---

[4] Lilian Edwards, "Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective." *European Data Protection Law Review (EDPL)* 2, no. 1 (2016): 28 (33),
[5] CAO, "The 5th Science and Technology Basic Plan", Council for Science, Technology and Innovation Cabinet Office, Government of Japan, 2015,
The Science and Technology Basic Plan is a five-year plan mandated by the 1995 Science and Technology Basic Law. It maps the science and technology policy of the Japanese government. The first plan was presented in 1996 and the 6th version was published in 2021.
[6] Andrew DeWit, "Japan's Rollout of Smart Cities: What Role for the Citizens?" *The Asia-Pacific Journal* , no. 24 (2013), accessed October 22, 2023, https://apjjf.org/2014/11/24/Andrew-DeWit/4131/article.html.
[7] Tim Hornyak, "Why Japan is Building Smart Cities from Scratch," *Nature Spotlight* 608, no. 32 (2022), accessed October 22, 2023, https://www.nature.com/articles/d41586-022-02218-5.
[8] "Toyota to Build Prototype City of the Future", Toyota, January 7, 2020, accessed October 22, 2023, https://global.toyota/en/newsroom/corporate/31171023.html.

Gassman, *Böhm*, and *Palmié* indicate that as a city's products, processes, and services become intelligent, autonomous, interconnected, and integrated to facilitate ecological and social improvements, they obtain a digital shadow. Initially neutral and without purpose, this shadow is imbued with meaning and significance by its surroundings.[9]

Traditional analysis methods cannot keep up with processing the sheer amount of data that makes up a city's digital shadow. Hence, to mold that initially purposeless shadow, smart cities utilize algorithms in the form of big data analytics. AI allows for data-based, also called evidence-based, decision-making using a city's digital shadow.

This data-driven approach to decision-making receives multiple names: autonomous or automated decision-making, AI-driven decision-making, AI-integrated decision-making support systems, or intelligent decision-making support systems, to name a few. This chapter will refer to these systems as autonomous decision-making (ADM). ADM is already being used to improve public services in areas such as finance, healthcare, marketing, commerce, command and control, and cybersecurity.[10] In addition, commercially available applications also utilize AI to guide users in multiple areas: from autonomous vehicles to map applications that track traffic in real-time to applications suggesting exercise programs or diets based on user preferences.

In contrast to other issues, such as privacy or data protection, rules on AI use are still in their infancy. Thus, the various legal aspects of its use are actively being discussed at the academic and lawmaking level. This chapter presents a brief overview of Japanese efforts in the matter. It begins with an introduction to the concept of smart cities. Next, it explores issues regarding the risk of utilizing AI, followed by an exposition of current regulatory challenges and efforts. Finally, the chapter concludes with Japan's efforts to regulate AI in Society 5.0 and a brief discussion on a recent AI profiling scandal.

I. Smart cities

While the origins of the term smart city can be traced to the 1990s, currently, there is no single accepted definition. The UN's United for Smart Sustainable Cities (U4SSC) initiative focuses on the sustainability aspect of smart cities, which it defines as:

> *"A smart sustainable city is an innovative city that uses information and communication technologies (ICTs) and other means to improve quality of life, efficiency of urban operation*

[9] Oliver Gassmann, Jonas Böhm and Maximilian Palmié, *Smart Cities: Introducing Digital Innovation to Cities.* (Bingley: Emerald Publishing, 2019), 28.

[10] Gloria Phillips-Wren,"AI Tools in Decision Making Support Systems: A Review," *International Journal of Artificial Intelligence 20*, no.10(2012): 1 (1).

> *and services, and competitiveness, while ensuring that it meets the needs of present and future generations with respect to economic, social, environmental as well as cultural aspects."*[11]

The European Commission provides a broader definition of smart cities:

> *"A smart city is a place where traditional networks and services are made more efficient with the use of digital solutions for the benefit of its inhabitants and business.*
>
> *A smart city goes beyond the use of digital technologies for better resource use and less emissions. It means smarter urban transport networks, upgraded water supply and waste disposal facilities and more efficient ways to light and heat buildings. It also means a more interactive and responsive city administration, safer public spaces and meeting the needs of an ageing population."*

Even at the academic level, there is no consensus on a smart city definition. *Albino*, *Berardi*, and *Dangelico* list over 20 definitions dating back to 2000.[12] *Hollands* highlights the difficulties of portraying a city as smart, from the terms used to describe them, the marketing hype, and the uncritical stance towards the urban development model they adopt.[13] A solution, as presented by *Kitchin*, is to approach the issue via classification. For example, by dividing smart cities' definitions into two camps. The first emphasizes the pervasiveness of digitally instrumented devices in the urban environment to monitor, manage and regulate city flows. The second camp prioritizes the knowledge economy within a region where the economy and governance are guided by innovation, creativity, and entrepreneurship.[14]

As the integration of digital devices has become the norm, classification efforts have shifted to the specific goals of a city. For example, *Alexopolous et al.* report that most initiatives focus on any combination of transportation, environment, tourism, health, waste management & water resources, energy-sustainable development, ICT infrastructure, economic development, security, and e-government goals.[15] Indeed, while most publications still tout the perceived benefits of smart cities, there seems to be a trend to focus on specific achievable goals. A possible reason is that most

---

[11] UNECE, "Guidelines for the Development of a Smart Sustainable City Action Plan". UNECE, 2018, https://unece.org/DAM/hlm/documents/Publications/Guidelines_for_SSC_City_Action_Plan.pdf

[12] Vito Albino, Umberto Berardi, and Rosa Maria Dangelico, "Smart Cities: Definitions, Dimensions, Performance, and Initiatives," *Journal of Urban Technology* 22, no. 1 (2015) 3 (6-8), doi: 10.1080/10630732.2014.942092.

[13] Robert G. Hollands, "Will the real smart city please stand up?," *City* 12, no. 3 (2008): 303 (305-307), doi: 10.1080/13604810802479126.

[14] Rob Kitchin. "The Real-time city? Big data and smart urbanism," *GeoJournal* 79, no. 1 (2014): 1 (1-2), doi: 10.1007/s10708-013-9516-8.

[15] Charalampos Alexopoulos, Gabriela Viale Pereira, Yannis Charalabidis, and Lorenzo Madrid, "A Taxonomy of Smart Cities Initiatives." *ICEGOV* '19: 281 (285), doi: 10.1145/3326365.3326402.

countries in the position to implement these initiatives have already gone through adapting or building the necessary infrastructure and can, therefore, shift their efforts to specific goals.

However, smart cities are not without their critics. *Shelton* and *Lodato* point out that even though they masquerade as a public good, these cities are essentially indistinguishable from earlier iterations of neoliberal urbanism. Instead of providing a cure-all to lagging cities, they reinscribe substantial urban social and spatial inequalities by privileging free market, technology-centric, and expert-driven forms of urban planning and governance.[16] *Kitchin* warns that the promotion of smart cities by several of the world's largest software services and hardware companies might lead to the corporatization of city governance and technological lock-in.[17]

From a legal perspective, the increased participation of the private sector in the administration of cities presents a new set of challenges. Smart city initiatives follow each country's constitutional principles; those that adhere to a western political ideology usually highlight individual rights, while those based on eastern political thought tend to focus more on social order and control.[18] There are numerous examples of countries limiting internet access in times of social turmoil. Hence, the issue of whether traditional constitutional and human rights protection are suited for a smart society is still undecided.

Additionally, the vast amount of data collected also raises privacy and data protection concerns, far too many to tackle in this chapter. *Edwards*[19] and *Van Zoonen*[20] describe the privacy challenges in a smart society. These range from data collection, protection, and combined use of data points to identify individuals to the difficulties of obtaining consent. Regarding this last point, *Finch* and *Tene* note that public services have captive populations who cannot opt-out of information collection without paying a steep price for safety, convenience, and quality of life. Thus, under this urban model, citizens might not have a choice which raises questions as to whether they can consent to provide it in the first place.[21]

---

[16] Taylor Shelton and Thomas Lodato,"Actually Existing Smart Citizens Expertise and (Non)participation in the Making of the Smart City." *City* 23 (2019): 35 (35), doi: 10.1080/13604813.2019.1575115.

[17] Rob Kitchin, (n 14) 10.

[18] Ruben Rodriguez Samudio, "Desafíos constitucionales de las ciudades inteligentes," *CES Derecho* 12, no. 2 (2021): 3 (9-11), doi: 10.21615/cesder.6174.

[19] Lilian Edwards, (n 4) 28-58.

[20] Liesbet Van Zoonen, "Privacy concerns in smart cities," *Government Information Quarterly* 33 (2016): 472-80.

[21] Kelsey Finch and Omer Tene, "Welcome to the Metropticon: Protecting Privacy in a Hyperconnected Town." *Fordham Urban Law Journal* 41 (2014): 1581 (1596).

II. Smart cities' citizens

Beyond an academic or industry definition, there is also the question of who makes up smart cities. *Sadowski* and *Bendor* note that the sensors, networks, and algorithms associated with a smart city can be deployed in other settings and contexts. These technologies are neither strictly technical nor social, becoming smart only by their link to the idea of a smart city and the symbolism it embodies.[22]

The citizens, governments, and corporations share this imaginary, and this relationship can be observed in how smart cities are designed and built.
Using a Triple-Helix model, *Leydesdorff* and *Deakin* posit that

> "*[C]ities can be considered as densities in networks among at least these three relevant dynamics: that is, in the intellectual capital of universities, the industry wealth creation, and their participation in the democratic government which forms the rule of law in civil society."*[23]

While Leydesdorff and Deakin approach the city as the convergence of collaboration efforts between government, academia, and industry, their model does not consider direct citizen participation. To address this, other authors[24] propose a modified Quadruple-Helix model that includes civil society in developing smart cities. Recently, a Penta-Helix (Quintuple-Helix) model has appeared. This model provides for the public and private sectors, academia, civic society, and social entrepreneurs. [25] In addition, more and more governments have begun including citizen participation in the design of their policies. As a result, citizens are no longer mere users but both drivers and subjects of change. This increased emphasis on citizen participation responds to the mounting criticism of smart cities.[26]

Two of the four models *Niarios* described in his taxonomy smart cities center on citizen participation: resilient smart cities and commons-based smart cities.[27] Resilient smart cities adopt a bottom-up approach to foster new participatory planning and governance forms. Under this model, communities have ownership and control of the infrastructure, and users can interact privately within a local network and avoid sharing details beyond it. On the other hand, commons-based smart cities are

---

[22] Jathan Sadowski and Roy Bendor, "Selling Smartness: Corporate Narratives and the Smart City as a Sociotechnical Imaginary." *Science, Technology, & Human Values* 4, no. 3 (2018): 540 (541), doi: 10.1177/0162243918806061.
[23] Leydesdorff and Deakin, "The Triple-Helix Model of Smart Cities: A Neo-Evolutionary Perspective." *Journal of Urban Technology* 18, no. 2 (2011): 53 (63), doi:10.1080/10630732.2011.601111.
[24] Patrizia Lombardi, et al., "Modelling the Smart City Performance," *The European Journal of Social Science* 25, no. 2 (2012): 137-149, doi: 10.1080/13511610.2012.660325.
[25] Igor Calzada, «From Smart Cities to Experimental Cities?» In Co-Designing Economies in Transition, ed. Vincenzo Giorgino and Zack Walsh (Cham, Switzerland: Palgrave Macmillan, 2018)
[26] Taylor Shelton and Thomas Lodato, "Actually Existing Smart Citizens," *City* (2019) 35 (36), doi: 10.1080/13604813.2019.1575115.
[27] Vasilis Niaros, "Introducing a Taxonomy of the "Smart City": Towards a Commons-Oriented Approach?" *tripleC* 14, no. 1 (2016): 51 (57-58), doi: 10.31269/triplec.v14i1.718.

characterized by wide citizen engagement in the design and implementation of technological infrastructures. Furthermore, they promote continuous innovation and knowledge diffusion on a global scale via an ongoing circulation of the commons.

Resilient smart cities and commons-based smart cities are indeed enticing. However, the technical knowledge required to participate in these cities is by no means low. There is the risk that compulsory technological engagement in city services could alienate a significant part of the population, particularly in countries with aging populations, such as Japan. Even without the obstacle of age, economic differences might exacerbate inequality by limiting access to those services that require digital devices. In addition, not everyone can achieve the necessary level to collaborate constructively on specific topics, such as AI.

A technologically developed country does not necessarily translate into digital literate citizenship. For example, in a study on perceived surveillance, *Jameson*, *Ritchner,* and *Taylor* report how the citizens of Amsterdam participants often spoke about "the government" as a monolithic entity. In reality, there is no data department of the city municipality, and in some cases, the municipality does not have the technical capabilities and so must outsource some of the analytics.[28] Thus, citizens' expectations might not align with the services they receive, regarding results and how said services operate.

In the case of Japan, a poll from the Ministry of Internal Affairs and Communications reveals that over 89 % of polled individuals have smartphones. Still, only 48 % of respondents have laptops, and a lower percentage of people (26.5 %) use desktop computers.[29] Furthermore, the country ranks 62nd out of 63 countries on digital literacy[30], and there is a marked shortage of skilled labor, with over 89 % of businesses expressing a need for IT-experienced employees.[31]

B. AI in smart cities

Big data analytics is one of the pillars that support smart cities. As collected data increases, the algorithms required to process and use said data become more complex, so it is no longer possible for human actors to perform these tasks. Instead,

[28] Shazade Jameson, Christine Richter, and Linnet Taylor, "People's strategies for perceived surveillance in Amsterdam Smart City," *Urban Geography* 40, no. 10 (2019): 1467 (1474-1475), doi: 10.1080/02723638.2019.1614369.

[29] MIC. Reiwa 2 Nendo Uizu Korona ni okeru Dejitaru Katsuyo no Jitsumu to Riyosha Ishi no Henk ani kansuru Chosa Kenkyu no Ukeoi – Hokokusho-. (Tokyo: Ministry of Internal Affairs and Communications, 2021) (in Japanese).

[30] IMD. "World Digital Competitiveness Ranking". IMD.International Institute for Management Development, 2022. https://www.imd.org/centers/wcc/world-competitiveness-center/rankings/world-digital-competitiveness-ranking/#:~:text=The%20yearly%20ranking%20%E2%80%93%20one%20of,the%20first%20time%20since%202017.

[31] MIC. Dejitaru de Sasaeru Kurashi to Keizai. (Tokyo: Ministry of Internal Affairs and Communications, 2021) (in Japanese).

AI makes sense of the collected data and provides fast, accurate results that can be applied immediately to run public services or within the private sector.

Same as with smart cities, defining AI is a challenging endeavor. As a guide, we turn to *Russell* and *Norvigs* classification.[32] They classify AI definitions based on whether they describe thought processes, reasoning, or behavior, and how they fare against human or ideal (rational) performance. Hence, AI can be defined based on whether they act "humanly" or "rationally", or whether they think "humanly" or "rationally".

There is also a philosophical definition of AI that classifies AI into weak AI, machines that *simulate* or *act* as if they were intelligent, and strong AI, machines that can *actually think*.[33] Moreover, these terms have a narrower meaning when referring to practical application. For example, IBM defines them as follows:

> *"Weak AI, also known as narrow AI, focuses on performing a specific task, such as answering questions based on user input or playing chess. It can perform one type of task, but not both, whereas Strong AI can perform a variety of functions, eventually teaching itself to solve for new problems. Weak AI relies on human interference to define the parameters of its learning algorithms and to provide the relevant training data to ensure accuracy. While human input accelerates the growth phase of Strong AI, it is not required, and over time, it develops a human-like consciousness instead of simulating it, like Weak AI."*[34]

While weak AI has already been achieved and is used to solve multiple issues across various fields, the consensus is that strong AI still belongs to sci-fi. In a review of big data analytics in smart cities, *Soomro et al.* report that AI benefits cut across multiple domains such as transportation planning, urban planning, smart buildings, weather prediction, and analysis.

I. Smarter cities?

The vast quantities of data collected by smart cities make it unusable with traditional methods. *Gassman*, *Böhm*, and *Palmié* explain that modern data analytics through algorithms identify patterns and autonomously improve systems without human intervention.[35] Therefore, arguably, the role of AI in smart cities is effectively to make the city smart enough to use the collected data.

---

[32] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, third edition (NewJersey: Pearson, 2016), 1-2.
[33] John Searle, "Minds, brains, and programs," *Behavioral and Brain Sciences* 3, no. 3 (1980): 417-424, doi: 10.1017/S0140525X00005756.
[34] "Strong AI," IBM, August 31, 2020, accessed October 22, 2023, https://www.ibm.com/cloud/learn/strong-ai.
[35] Gassmann, Böhm and Palmié, (n 9) 273-274.

Soomoro et al. point out that smart cities function by sharing data and information among their various domains and that these domains have traditionally acted as information silos. However, there is an expectation for these barriers to be broken and for information to flow freely, with policymaking being an integrated and holistic process.[36] By combining the data gathered via IoT technologies, ADM can link these information silos to provide multiple solutions to achieve specific goals within the city.

As *Del Gamba* indicates, ADM is a comprehensive concept not limited to one type of decision-making process. In examining the possible application of the GDPR to ADM systems, the author classifies them into those with profiling and those without. The former refers to systems that reach a solution based on a deterministic firm rule that guides the decision-making process automatically, such as speeding fines imposed by speed cameras. On the other hand, ADM uses profiling techniques to collect personal data, process it to identify correlations and apply a model to identify present or future behavior.[37]

Even though *Del Gamba* approach to ADM focuses on GDPR obligations, this classification can be applied to all ADM systems within the city. Furthermore, profiling ADM is not limited to identifiable personal data; it also extends to decisions that might affect certain groups, either via policies directed at them or indirectly via policies that affect the areas where they reside or work.

As *O'Neil* observes, there is always the possibility that the mathematical models that fuel ADM are encoded with human prejudice, misunderstanding, and bias. She coined the term 'weapons of math destruction' to refer to these mathematical models and argues how, while promising efficiency and fairness, they distort higher education, drive up debt, spur mass incarceration, pummel the poor at nearly every juncture, and undermine democracy.[38] Likewise, *Eubanks* posits that the skyrocketing economic insecurity of the last decade has been accompanied by an equally rapid rise of sophisticated data-based technologies in public services, such as predictive algorithms, risk models, and automated eligibility systems. These new systems of digital poverty, as the author calls them, target the poor and the working class.[39]

---

[36] Kamran Soomro, Muhammad Nasir, Mumtaz Bhutta, and Zaheer Khan, "Smart City Big Data Analytics: An Advanced Review", *WIREs Data Mining and Knowledge Discovery* 9, no. 5 (2019). , doi: 10.1002/widm.1319.

[37] Giulia Del Gamba, «Machine Learning Decision-Making: When Algorithms Can Make Decisions According to the GDPR.» In *Law and Technology in a Global Digital Society Autonomous Systems, Big Data, IT Security and Legal Tech*, ed. George Borges and Christoph Sorge, (Springer, Cham, 2022), 76-77.

[38] Cathy O'Neil, *Weapons of Math Destruction* (New York: Crown, 2016).

[39] Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, (St. Martin's Press, 2018), 9-11.

Both authors call attention to a characteristic of ADM that, while evident, is not necessarily considered in their implementation: their complexity. In this sense, complexity is not limited to the required technical knowledge or the complex systems they link; it also refers to the complex social requirements and effects they produce either by design or by accident[40].

On the technical side, one criticism of ADM systems is that they are usually "black boxes", i.e., they produce accurate results without giving a detailed explanation or reasoning as to how they arrived at that solution. Moreover, because of the opaqueness of such decisions, it is difficult for people to assess whether they were discriminated against based on racial origin.[41]

Pasquale's book "The Black Box Society" is a grim reminder of the risks these systems entail. Specifically, he describes how companies race to keep their methods secret and how, even if they were willing to expose them, some sectors, such as banking or the Internet, pose a thought challenge because of their very nature.[42]

One proposed solution to the black box issue is explainable AI (XAI), programmed to explain its purpose and rationalize the decision process in a way the average user can understand.[43] *Longo et al.* note that there are multiple challenges surrounding XAI. First, in an ideal world, machine explanations and human understanding would be identical and congruent with the ground truth. However, in practice, the ground truth cannot always be fully defined in highly uncertain areas, such as medical diagnoses. Moreover, human models are often based on causality, which is particularly challenging as current machine learning follows pure correlation.[44]

The authors further indicate that current XAI methods highlight input–relevant parts that significantly contributed to a particular output or the most pertinent features of a training data set most influential to the model's accuracy. Unfortunately, XAI methods do not incorporate the notion of human models. They, therefore, do

---

[40] Barocas and Selbst identify five possible ways in which AI might accidentally create discrimination: 1. The way variables are defined; 2. The labeling of training data; 3. Issues with the data used as training data; 4. Issues with the attributes used to reach a decision or "feature selection"; and 5. When criteria that are genuinely relevant to make rational and well-informed decisions nevertheless serve as a reliable proxy for a class. Solon Barocas and Adrew Selbst. "Big Data's Disparate Impact" *California Law Review* 104, no.3 (2019): 671 (677-693), doi: 10.15779/Z38BG31.

[41] Frederik Zuiderveen Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, (Strasbourg: Council of Europe, Directorate General of Democracy, 2018), 15.

[42] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*,(Harvard University Press, 2015).

[43]Ana Carolina Borges, et al., «An Overview of Explainable Artificial Intelligence (XAI) from a Modern Perspective», In *Explainable Artificial Intelligence for Smart Cities*, ed. Mohamed Lahby, Utku Kose and Akash Kumar Bhoi(CRC Press, 2021), 4.

[44] Luca Longo, et al.., «Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions», In *Machine Learning and Knowledge Extraction,* ed. Holzinger A, Kieseberg P, Tjoa A, Weippl E (Dublin: Springer, Cham, 2020), 10.

not consider causability, defined as "the extent to which an explanation of a statement to a human expert achieves a specified level of causal understanding with effectiveness, efficiency, and satisfaction in a specified context of use and the particular contextual understanding capabilities of a human."[45]

Technical difficulties aside, ADM, powered either by traditional AI or XAI, expert-focused explainability does not solve the trust issue. In other words, even if XAI is developed to the point that human experts can understand and replicate the decision-making process, this would not benefit most citizens. Particularly in those cities that aim for a Penta-helix model, citizen participation will most likely be hindered if only those with advanced technical knowledge can provide input.

While most smart cities are developed under governmental auspices, their success depends on a large part of the citizens, particularly their willingness to adopt certain technologies. In the case of ADM, this means that once a certain usage threshold has been achieved, the technology can continue to grow and evolve, albeit it might do so with incomplete or biased data.

It could be argued that since the average citizen does not have a working knowledge of most city services, such as trains or energy grids, ADM need not be treated differently than other public services. However, this argument fails to convince for multiple reasons. First, while it is true that an average citizen could not perform complex tasks associated with city administration, most citizens would be able to observe, or at the very least have a clue, that these services are not operating correctly. By contrast, as *O'Neil, Eubanks*, and *Pasquale*' demonstrate, identifying issues in ADM systems, such as biases in the data or adverse outcomes, is a time-consuming process that most citizens would most likely not be able to perform.

Another issue has to do with the type of decision being taken. Decisions, as *Phillip-Wren* explicates, can be divided into structured, those with a clear answer; unstructured, where there is no agreed solution and depend on the decision maker; and semi-structured, which can be represented with analytical models or are based on data.[46] Although citizens may agree with most structured decisions, particularly those that do not rely on collecting personal data, the same cannot be said about unstructured and semi-structured decisions.

In the case of unstructured decisions, ADM systems might provide multiple options, all of them valid. However, since there is no agreement on a solution, it will depend on politics and the personal beliefs of decision-makers. In a sense, ADM systems provide the decision-maker with the tools to make a more informed decision, but at the same time, they also might bring questions of responsibility. For example, who is to blame if a particular policy, based on ADM recommendations,

[45] Luca Longo, et al., (n 44).
[46] Gloria Phillips-Wren, (no 10)1-2.

creates problems down the line? The decision-maker? The contractor? Or citizens who did not understand the system enough yet initially benefitted from it?

In privacy and data protection literature, there is the phenomenon known as privacy fatigue, defined as a sense of weariness toward privacy issues, in which individuals believe that there is no effective means of managing their personal information on the Internet.[47] Hence, it is not unreasonable to believe that a similar issue could arise regarding ADM systems, where decision-makers, be it public officials, corporations, or citizens, might decide to follow ADM suggestions without questioning them. For example, suppose a decision based on a model that provides what appears to be an adequate solution. Then, what incentives do decision-makers have to question a solution provided via ADM? In other words, ADM runs the risk of creating systems where human's sole role as decision-makers is to put a stamp of approval, regardless of whether they understand or agree with said solution.

II. Regulating AI

An immediate response to the challenges mentioned would be to regulate. However, and the same as any emerging technology, regulatory efforts also face several difficulties. Black and Murray argue current discourse on regulation is drawing us away from the law, or even traditional models of command and control or co-regulation and governance, towards soft self-regulation and codes of practice.[48]

They describe this as an ethical model that has seen the adoption of codes of practice for general AI and data-driven health and care technology, among others. However, the authors caution that ethical standards for such systemic risks are insufficient, particularly in assuming that risks are individualized and that the key to their management is the individual consumer's choices within the marketplace.

Scherer observes that traditional regulation methods are unsuited to manage the risk associated with intelligent and autonomous machines. Particularly, he argues that ex-ante regulation would be difficult because AI research and development may be discreet (requiring little physical infrastructure), discrete (different components of an AI system may be designed without conscious coordination), diffuse (dozens of individuals in widely dispersed geographic locations can participate in an AI project), and opaque (outside observers may not be able to detect potential harmful features of an AI system). Furthermore, the autonomous nature of AI creates issues of foreseeability and control that might render ex-post regulation ineffective, particularly if an AI system poses a catastrophic risk.[49]

---

[47] Hanbyul Choi, Jonghwa Park, and Yoonhyuk Jung, "The Role of Privacy Fatigue in Online Privacy Behavior." *Computers in Human Behavior* 81, (2018): 42 (42), doi: 10.1016/j.chb.2017.12.001.
[48] Julia Black, and Andrew Murray, "Regulating AI and Machine Learning: Setting the Regulatory Agenda," *European Journal of Law and Technology* 10, no. 3 (2019).
[49] Matthew U Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies," *Harvard Journal of Law & Technology* 29, no. 2 (2016): 354 (356-357).

Regardless of these challenges, the push for regulation has gained momentum, with many countries enacting laws or publishing guidelines that regulate AI. [50] For example, in 2020, the US passed the National Artificial Intelligence Act to coordinate efforts regarding AI research at the Federal Level. Likewise, the are multiple bills at the state level that aim to regulate or promote various aspects of AI. More recently, the White House published the Blueprint for an AI Bill of Rights, which sets forth five principles: safe and effective systems, algorithmic discrimination protections, data privacy, notice and explanation, and human alternatives, consideration, and fallback.[51]

The AI Bill of Rights establishes that Designers, developers, and deployers should take proactive measures to protect individuals, including equity assessments, use of representative data, protection against proxies for demographic features, accessibility, organizational oversight, independent evaluation, and plain language reporting. Furthermore, the human alternatives, consideration, and fallback principle requires that individuals have the option to opt out of automated systems in favor of a human alternative, where appropriate.

At the European level, the European Parliament's proposal for harmonized rules on artificial intelligence (Artificial Intelligence Act) aims for a comprehensive set of regulations concerning AI. The draft classifies AI depending on its objectives and uses into limited or minimal-risk, high-risk, and unacceptable-risk. The European Commission's website[52] explains minimal-risk AI encompasses most current applications, such as video games or spam filters. Limited-risk applications, such as chatbots, are subject to some transparency obligations that would allow the user to know they are interacting with an AI.

An AI is considered high-risk if it meets the following two criteria: 1. It is intended to be used as a safety component of a product or is itself a product; and 2. The product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product according to various harmonization legislations.

In addition, Articles 12, 13, and 14 establish record-keeping, transparency, and human oversight requirements, respectively. Thus, AI must create an automatic recording of events (logs) that ensure that AI functioning can be traced during its lifetime. Moreover, AI systems must be developed to ensure their operation is transparent enough to allow users to interpret their output and use it appropriately.

50 The OECD has a comprehensive database on various countries AI initiatives, "National AI policies and Strategies", OECD.AI Policy Observatory, OECD.AI, 2022, https://oecd.ai/en/dashboards/overview.

51 "Blueprint for an AI Bill of Rights", The White House, accessed October 22, 2023, https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

52 "Regulatory Framework Proposal on Artificial Intelligence," European Commission, accessed October 22, 2022, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

Lastly, AI systems must be designed and developed in a manner that will enable human oversight.

The draft also lists numerous examples of unacceptable-risk AI. For example, when used by public authorities or on their behalf, AI is considered unacceptable if used to evaluate or classify the trustworthiness of natural persons based on their social behavior or known or predicted personal or personality characteristics. Furthermore, AI is also prohibited if such classification leads to detrimental or unfavorable treatment unrelated to the context where the data originated or was collected or is unjustified or disproportionate to their social behavior or its gravity.

III. AI in Society 5.0

Japan has redoubled its efforts to implement AI. The 2017 Strategic Council for AI Technology's "Artificial Intelligence Technology Strategy" established a three-phase roadmap.

The goal is to fuse AI with other technologies to improve productivity, health, medical care, welfare, and mobility. Phase one focuses on the utilization and application of data-driven AI. Productivity-wise, this translates into mass customization, on-demand supply services, and smart factories using IoT and AI. In the healthcare industry, the roadmap aims to implement telemedicine, home medical care, AI-assisted examination, and smart operating rooms. Lastly, some mobility-related goals are to increase car-sharing services, expand GPS-related industries, diversify transportation services, and spread telecommuting in IT.

Phase two, which was supposed to begin around 2020 and should end from 2025 to 2030, emphasizes the public use of AI and data. Specific goals to increase productivity include implementing dynamic pricing, automatic replenishment services, house and home appliances powered by AI, eliminating energy waste, and automatic maintenance of machinery and equipment. Health-wise, some of this phase's objectives are complete medical checkups at home, constant health monitoring services, artificial organs, and even nanobots. Concerning mobility, phase two concentrates on expanding the sharing economy, autonomous transportation and delivery, level 4 autonomous cars, and privatization of travel time and space.

Lastly, phase three aims ecosystem built by connecting multiplying domains. This translates into a society "where innovative services and products are continuously developed" (productivity), "that enjoys healthful lifestyle and longevity" (health), and "that enables safe and free travel" (mobility).

So far, results have been mixed. Some goals, such as increasing car-sharing services, have indeed been achieved. The global pandemic also forced companies to adopt work from home, though it is unclear whether the trend will continue once the situation normalizes. It also helped to increase the use of digital payment services. By contrast, other areas, such as telemedicine, home medical care, and complete medical checkups at home, are still far from being a reality.

In addition, the 2019 report "Social Principles of Human-Centric AI" sets forth the principles that guide the role of AI in Society 5.0. These social principles are

based on three philosophies. First, respect for human dignity translates into a society that does not try to control human behavior through the pursuit of efficiency and convenience and where AI is used only as a tool. The second philosophy is diversity and inclusion, under which people with diverse backgrounds, values, and ways of thinking can pursue their own well-being. Last is creating a sustainable society where AI is used to resolve social disparities while tackling issues such as climate change.

Moreover, the report presents two categories of principles: social and R&D. The socials principles of AI, as detailed in the report, are as follows:

- Human-centric principle: AI should not infringe upon fundamental human rights. Furthermore, AI use should be an individual decision. Stakeholders involved in the development, delivery, and utilization of AI should, depending on the nature of the problem, be responsible for the consequences of AI utilization.
- Education/literacy principle: policymakers and managers of businesses involved in AI must accurately understand AI. By contrast, AI users should have a general understanding. Therefore, educational environments should foster this knowledge by providing a system that allows anyone to understand the basics of AI, mathematics, and data science.
- Privacy protection principle: when utilizing AI, a higher level of discretion may be required than the mere handling of personal data according to the data's level of importance and sensitivity. This principle is a restatement of current privacy laws adapted to AI applications.
- Security principle: Society should always be aware of the balance between the benefits and risks that AI entails.
- Fair competition principle: AI should create a competitive environment that promotes new business. In addition, a country or company's dominant position regarding AI should not determine sovereignty, data collection, or competition practices.
- Fairness, accountability, and transparency principle: AI uses should also translate into a fair and transparent decision-making process with appropriate accountability for the results. Moreover, appropriate explanations should be given on a case-by-case basis depending on the application of AI with adequate opportunities for an open dialogue, as required, regarding the use, adoption, and operation of AI.
- Innovation principle: Society 5.0 should aim for continuous innovation that transcends boundaries, promoting total globalization, diversification, and industry-academia-government cooperation.

Regulation-wise, Japan falls behind other jurisdictions. For one, there is no statutory definition of AI. However, the government's AI Strategy defines AI as a system that realizes an intelligent function. In 2021, the Expert Group on How AI Principles should be implemented published the "AI Governance in Japan Ver. 1.1" report. The expert group adopts an industry-centered approach and recommends

legally non-binding corporate governance guidelines as the most desirable solution. Specifically, the report argues that:

> *"In developing guidelines, not only companies that use AI but also a wide range of stakeholders, including users, engineers, academics, and law/audit experts, should engage in the discussion. It is desirable that the government functions as a facilitator in the discussion and objectively evaluates whether companies satisfy the guidelines developed, thereby enhancing society's trust in the companies that meet the guidelines."*

Specifically, the report calls for guidelines to avoid using standards based on a specific level of experience and one-size-fits-all application to all companies. In addition, guidelines should support the improvement of AI risk management, and function as a benchmark of the trustworthiness of AI systems in inter-company transactions. Ideally, these guidelines should also include helpful practices for companies that have just started using AI, and facilitate the provision of explanations to consumers, etc.

Moreover, the expert group considers that legally binding horizontal requirements for AI systems are unnecessary and that efforts should focus on customer literacy instead. They also oppose mandatory AI-based technology regulation, and even when it is needed, its scope should be limited to prevent them from impacting unintended areas. The report goes one step beyond and argues that it is desirable to respect rulemaking in the respective sectors by making the most of the existing concept of regulations and design philosophy" in some specific industries, such as the automotive and healthcare sectors.

Interestingly, the expert group does not tackle the issue of education/literacy and the impact it might have in protecting users. Instead, it merely points out that the Consumer Affairs Agency released a handbook titled "Handbook on Use of AI-Keys to Effectively Using AI" in 2021. The expert group report explicitly states that one of the reasons for deeming legally binding horizontal regulation unnecessary is the "direction of improvement of literacy through the Handbook on Use of AI."

Recently, Japan faced its first scandal regarding AI profiling. In 2019, the Rikunabi Data Scandal in which a recruitment company sold student's data to prospective employers. Fudo, Arai, and Ema[53] explicate that the Japanese custom of "simultaneous recruitment of new graduates" provides the necessary social background for the scandal to occur. Since most new graduates engage in job hunting simultaneously, some may receive informal job offers from several companies. Therefore, there is always the possibility that some of them might reject an offer,

[53] Kudo Fumiko, Hiromi Arai, and Arisa Ema, "Ethical Issues Regarding the Use of AI Profiling Services for Recruiting: The Japanese Rikunabi Data Scandal.". Arvix, 2020. https://arxiv.org/ftp/arxiv/papers/2005/2005.08663.pdf#:~:text=Ethical%20Concerns%20on%20Power%20Structure&text=The%20incident%20illustrates%20that%20Recruit,should%20recognize%20the%20power%20structure.

thus putting the rejected employers at a disadvantage since they might be unable to fill alternative employees.

The recruitment company operates the Rikunabi job-placement website, which matches employers with job seekers. This website requires job seekers to access its database. In 2018, the company unveiled its "Rikunabi DMP follow" service, which collected and analyzed demographic information and cookies to calculate the probability of students declining informal job offers for a specific company. The algorithms used for profiling were unique for each company and were derived from comparing the users who accepted the offer and those who declined the previous year. The specific algorithms are not published[54].

The main issue in the Rikunabi Data Scandal was not one of profiling via AI, instead it was one of privacy. In 2019, the Rikunabi DMP follow service calculated the score for 74,878 users. At the time, cookies or machine-generated identifiers did not fall within the purview of the Japanese Personal Data Protection Act if they could not be used to identify a person. Since the company was only with data disconnected from names and other personal identifiers, systematically unready to collate additional information to identify individuals, it was not collecting, processing, and providing any personal data and did not need to get user consent to calculate and deliver algorithmic scores to client companies.[55] Moreover, due to a change in the privacy policy, 7983 users were not informed that their information could be provided to companies. This scandal resulted in amendments to the Personal Data Protection Act that established stricter rules on using personal data and its transfer to third parties.

While the social impact of the scandal was primarily felt in privacy circles, since this was the first case concerning "artificial intelligence," it prompted discussions on its use. Due to this publicity, some engineers are concerned that the idea that artificial intelligence is used in recruitment might be linked to a negative image in the public's eye.[56]

C. Conclusions

While discussions on the rights of robots and AI consciousness are still the domain of Sci-fi novels, ADM systems are already being used in government, industry, and entertainment. As a result, there is a renewed emphasis on AI regulation. Japan's Society 5.0 initiative plans to make extensive use of AI technologies.

---

[54] Fumiko, Arai, and Ema, (n 53).

[55] Hinako Sugiyama, Katitza Rodriguez, and Bennet Cyphers. "Japan's Rikunabi Scandal Shows the Dangers of Privacy Law Loopholes," Electronic Frontier Foundation, May 12, 2021, accessed October 22, 2023. https://www.eff.org/deeplinks/2021/05/japans-rikunabi-scandal-shows-dangers-privacy-law-loopholes.

[56] Fumiko, Arai and Ema, (n 53).

In contrast to the EU's Artificial Intelligence Act, Japan has yet to develop or present clear rules concerning ADM systems. Instead, the Japanese approach to regulation, or at the very least the Expert Group's recommendations, are based on an industry-centric regulatory model while trusting users to develop the required knowledge on AI to be able to judge by themselves.

However, this approach does not consider the current reality of the Japanese population's digital literacy. While technologically advanced, Japan has a relatively low level of digital literacy compared to similar countries, both at the general population level and in the IT field. Even though the literacy principle calls for a focus on AI-related subjects and the AI Strategy sets forth specific goals, in practice, whether these goals are achieved does not depend on a single government policy or strategy. Instead, AI use is driven by the cost-benefit analysis companies and individuals make when deciding whether to use a particular application or software. This analysis is based on economic costs, ease of use, results, etc.

Japanese society has a long tradition of seamlessly incorporating technologies into daily activities, and there is no reason to believe that AI will be any different. Nevertheless, a seamless implementation does not equal understanding. If the response, or lack thereof, to Rikunabi scandal is any indication, the Japanese government and the population at large still do not grasp the risks associated with ADM systems.

While still in the drafting stage, if the EU's Artificial Intelligence Act requires GDPR l of adequacy, the current Japanese approach, based on a lack of legally binding rules coupled with low digital literacy and an industry-focused approach to regulation, could lead to the two systems being incompatible.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP 22K13274.