



Research Centre for  
Sustainable Hong Kong

香港城市大學  
City University of Hong Kong

# ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE IN HONG KONG

REPORT



Research Centre for Sustainable Hong Kong (CSHK)

City University of Hong Kong

Chun-kit CHEUK, Ho-mun CHAN, Victor Tin-yau HUNG, Linda Chelan LI, Fen LIN, and  
Viktor TUZOV

**April 2023**

## About the Authors and Acknowledgment

---

**Mr. Chun-kit Cheuk** is a team member and PhD candidate of the research project at City University of Hong Kong. He drafted and revised the report with the input from the other authors.

**Dr. Ho-mun Chan** is the Principal Investigator of the research project and an Associate Professor of Philosophy in the Department of Public and International Affairs at City University of Hong Kong.

**Dr. Victor Tin-yau Hung** is a co-investigator of the research project and an Adjunct Associate Professor in the Faculty of Business and Economics at the University of Hong Kong.

**Prof. Linda Chelan Li** is a co-investigator of the project, Professor of Political Science, Department of Public and International Affairs, and Director of the Research Centre for Sustainable Hong Kong, City University of Hong Kong.

**Dr. Fen LIN** is a co-investigator of the project, Associate Professor of Media and Communication, Department of Media and Communication, City University of Hong Kong.

**Mr. Viktor Tuzov** is a PhD candidate at City University of Hong Kong.

## Acknowledgment

This report is the result of a collaborative effort by the authors under the project "*AI Governance: Ethics and Policy Challenges in Hong Kong*".

We would like to thank the interviewees and the Centre for Communication and Public Opinion Surveys (CCPOS) for their contributions to this report. We also want to express our gratitude to the Research Centre for Sustainable Hong Kong (CSHK) for their support in conducting the research. We are grateful for the funding support from our university through the Applied Research Grant (ARG) from 2019 to 2022, which made this research possible. Lastly, thanks are due to the Institute of Public Accountants (Hong Kong Office) and the Institute of Certified Management Accountants (Hong Kong Branch) for being supporting organizations of this project.

## About CSHK

Established in June 2017 by a cross-disciplinary research team, the Research Centre for Sustainable Hong Kong (CSHK) is an Applied Strategic Development Centre of City University of Hong Kong (CityU). CSHK conducts impactful applied research with the mission of facilitating and enhancing collaborations among academia, industry, professional service sectors, the community, and the government for sustainable development in Hong Kong and the region. For more information about CSHK, please visit [www.cityu.edu.hk/cshk](http://www.cityu.edu.hk/cshk). For enquiries, please email us at [sushkhub@cityu.edu.hk](mailto:sushkhub@cityu.edu.hk).

## How to cite this report

Cheuk, C. K., Chan, H. M., Hung, V. T. Y., Li, L. C., Lin, F., & Tuzov, V. (2023, April). Ethics and governance of artificial intelligence in Hong Kong. Research Centre for Sustainable Hong Kong.

# Content

<b>Executive Summary</b> .....	1
<b>Introduction</b> .....	5
<b>1.1 Structure of the report</b> .....	5
<b>1.2 Background</b> .....	6
<b>1.3 Research objectives</b> .....	7
<b>1.4 Working definition of terms</b> .....	8
<b>1.5 Literature review</b> .....	9
2.2.1 Literature research (I): Global perception of artificial intelligence technology .	9
2.2.2 Literature research (II): Global perceptions of AI technology and the ethical values of AI .....	13
2.2.3 Literature research (III): A preliminary summary of research directions.....	14
<b>1.6 Research methods</b> .....	15
2.2.4 Semi-structured interviews .....	15
2.2.5 Questionnaires.....	15
<b>Semi-structured interviews</b> .....	18
<b>2.1 Smart cities and AI applications (that arouse areas of concern)</b> .....	18
2.2.6 Background of the interview .....	18
2.2.7 Demand-driven smart cities .....	19
2.2.8 Understanding of intelligent applications.....	22
2.1.4 What people care about and what transparency means .....	24
2.1.5 Summary: Areas of concern.....	25
<b>2.2 Appropriate data governance framework</b> .....	26
2.2.1 Interpretation and education: Differences in understanding and expectations of experts .....	26
2.2.2 Hong Kong’s positioning and characteristics, and the role of the government	28
2.2.3 Summary: Views on governance .....	29
<b>Ethical values and principles in artificial intelligence: Questionnaire results</b>	30
<b>3.1 High ranking of universally endorsed ethical values</b> .....	30
<b>3.2 The clustering of ethical values under trade-offs</b> .....	30
<b>3.3 Value trade-offs contextualized under various technology types and scenarios</b>	31
<b>3.4 Characteristics of acceptance and response to technology types and scenarios</b>	33
<b>3.5 Governance: Whose views should be prioritized?</b> .....	34
<b>4. Outcome discussion and policy advice</b> .....	36
<b>4.1 Contextualization does matter</b> .....	36
<b>4.2 Framework on the application-scenario type and cultural differences</b> .....	37

<b>4.3 Explaining distrust towards enterprises.....</b>	<b>38</b>
<b>4.4 Meaning and strategies of educational promotion and communication</b>	<b>40</b>
<b>4.5 Conclusion.....</b>	<b>42</b>
<b>References.....</b>	<b>43</b>
<b>Appendices.....</b>	<b>45</b>
<b>Appendix A: Interview Protocol Outline .....</b>	<b>46</b>
<b>Appendix C: Questionnaire 調查問卷 .....</b>	<b>49</b>

## Executive Summary

- Due to increasing public attention to intelligent technology and its wide application in many practical aspects of our life, governments and technology enterprises from all over the world have laid down various versions of ethical principles on artificial intelligence (AI). More than 100 sets of such published ethical principles have been identified.
- After institutes and organizations at all levels have released documents declaring sets of ethical principles, textual analysis has been conducted on these documents, with the results of such research being published since 2019. The analysis has found a considerable degree of consensus among global institutions to uphold ethical values in the application of AI. However, by and large, these commitments have not been put into practice.
- Previous research reveals that our perceptions of AI are influenced by general demographics, such as gender, age, education, religious affiliation, and income, and that different countries have different ethical concerns.

Based on our **interviews**, it was found that our respondents held the following views:

### **Smart cities and AI applications (that arouse areas of concern)**

- The reason why some AI technologies have not been successfully implemented is not due to technical defects, but rather to other reasons, such as users' distrust of technology.
- Development of smart cities and AI projects have to be based on real substantive needs rather than the ambitious advancements of the technology providers.
- A common misconception about AI identified by respondents with an ICT or I&T background is that AI technology must rely on big data.
- A phenomenon of asymmetry: If a technology, such as Google map, is function-wise deeply entrenched in the daily life of a user without being aware of its introduction at the beginning, the part of the technology that involves AI will often not be morally controversial, but users become more skeptical when they are aware of the introduction of a new AI technology at the early stage.

### **Appropriate data governance framework**

- Good technology literacy will reduce the backlash of skepticism and instability, and reduce the obstacles faced by new technology projects.
- Technology providers and managers need to pay special attention to communication imbalances. Society tends to have a relatively vague understanding of the advantages

of technology, but a more specific one when it comes to disadvantages.

- Our respondents' answers echo certain ethical concerns expressed worldwide. Ethical values such as security, privacy, and fairness were mentioned by our interviewees, and the issue of transparency has attracted special attention and discussion. This issue does not only refer to the understanding of technology, but also includes the concerns of other values (e.g., privacy), which is induced by the lack of understanding.
- Expert-led communication is believed to be the core of governance. Technology experts should work together to lay down some ethical demands and establish consensus with the general public.
- From the perspective of the enterprises, they hoped that Hong Kong can have a common set of basic standards, with the government playing an important role of grounding these standards.

### Our **telephone survey** revealed the following:

- 500 Hong Kong residents were randomly phoned (landline and mobile) and asked to express their views on ethical values/principles and AI on three levels:
  - Decontextualized value ranking without trade-offs
  - Decontextualized value trade-offs
  - Scenario-application based value trade-offs

For the third level, we asked also whether they accept each of these AI applications and whose views should be prioritized in these scenarios.

### In **decontextualized value ranking**...

- The subjects were not alerted that the values in the list were often in conflict with each other in some situations. A general ranking of the list of ethical values is shown as below, with "Privacy" being the highest and "Unbiased" being the lowest:
  - Ranking of values: Privacy > Robustness > Individual Freedom > Transparency > Unbiased

### In **decontextualized value trade-offs**...

- In another set of questions, we asked our subjects to select three out of five ethical values while abandoning the other two and rank them accordingly – hinting at a trade-off between values. The ranking result did not change much.
- However, we observed a joint preference of "Individual Freedom", "Privacy", and "Transparency". Another set of preferences was "Robustness" and "Unbiased". The questionnaire provided another set of questions asking subjects to make trade-offs between values with specific AI application scenarios given. Unlike the decontextualized questions, there was a shift in the ranking of ethical values in AI with

more respondents believing that system “Transparency” is more important.

- Ranking of values: Transparency > Individual Freedom > Effectiveness (or “Robustness”)

### **In scenario-application based value trade-offs...**

- Respondents had divided opinions on the “Health Code System” and the “Fraud Detection System”: About half of the respondents believed that it should be deployed, while the other half believed that it should not. With regard to “Driverless Autonomous Vehicles”, roughly two-thirds of the respondents expressed opposition to the implementation of such systems, showing a more evident opinion inclination.

### **Acceptance and response to technology types and scenarios**

- The types and scenarios of AI application, and demographic factor affected the respondents’ perception towards AI applications.
- The responses to the “Health Code System” and the “Fraud Detection System” were quite consistent, with approximately 80% of the respondents expressing the same opinion on both types of AI applications: When they thought they would adopt one of the applications, there was a good chance that they would adopt the other at the same time, and vice versa. Driverless autonomous vehicle was considered as another type of AI technology and received different responses.

### **Whose views should be prioritized**

- 50-65% of the respondents who chose “difficult to choose” in the value trade-off question still tended to believe that “affected individuals” are the best persons to address these trade-off situations.
- On the whole, our respondents’ ranking order of who should be the best persons to address these trade-off situations is as follows, with “Affected Citizens” being the highest and “Private Businesses” the lowest:
  - Affected Citizens > Relevant Public Bodies > Third-Party Professionals >> Private Businesses (the least selected category with less than 2% of the respondents choosing)

### **Discussion and policy advice:**

- For those who want to foster AI ethics, including but not limited to policymakers and technologists, context makes difference – instead of simply listing out abstract ethical values or principles, focal points should be highlighted and specific context or scenarios of AI applications should always be provided.
- Special attention should be paid on catering to cultural and demographic diversity in the establishment of application-specific or sector-specific norms.



- The implementation of regulations should be incremental and according to the actual situation in Hong Kong – in both design and execution stages. Policymakers can make reference to the idea of “relevant markets” in the latest Competition Ordinance in Hong Kong.
- Agencies should be authorized to research, review, coordinate, and enforce standards and norms on AI applications. The feasibility of including the Application Impact Assessment in ESG (Environmental, Social, and Governance) reporting should be considered to allow for public scrutiny. Taking an incremental approach with constant reviews instead of setting up a rigid framework is advisable to ensure there is room for enterprises and individuals to explore, develop, and adapt to the market-oriented standards or performance pledge for AI applications on a voluntary basis.
- Support is necessary for research from third-party professionals, including statutory bodies and universities (which are independent from the government) to encourage independent research and offer empirically based policy advice, providing perspectives different from rigid, execution-oriented routine.

## Introduction

With the defeat of Korea's world Go champion by AlphaGo's DeepMind in 2016, there has been a growing interest in the application and potential impact of artificial intelligence (AI). The more recent big leap forward is ChatGPT of OpenAI, introduced at the end of 2022. The ChatGPT stunned the world with its power, leaving no doubt that AI will transform the way we live, conduct businesses, and govern.

As smart technologies, devices, and applications continue to mature, governments, private enterprises, and research institutions are recognizing their potential and grappling with the possible ethical and governance issues they raise. In 2017, the Government of the Hong Kong Special Administrative Region published its first Smart City Blueprint, followed by a 2.0 version in 2020. The blueprint serves as policy guidance and aims to strengthen related infrastructure, development, and talent training in artificial intelligence technology, with the hope of fully realizing its potential. Although issues relating to the ethics of technology, governance, and legislation are still in their early stages of formulation, researchers from the Research Centre for Sustainable Hong Kong (CSHK) recognize their importance in promoting a technology- and innovation-friendly environment and have worked on a research project entitled "AI Governance: Ethics and Policy Challenges in Hong Kong" since 2019. This report documents the findings from this research.

We hope that this research can inspire further discussion and more in-depth investigation in this area in future.

### 1.1 Structure of the report

The structure of this report is as follows:

- The introduction (this chapter) describes the background, research objectives and basic literature of this research project
- Chapters 2 to 3 present the key findings from interviews and survey
- Chapter 4 discusses the findings in conjunction with the analysis of foreign research and interviews to explore possible implications and provide policy recommendations

## 1.2 Background

### Origin of AI ethics and governance

The term "artificial intelligence" was coined about 70 years ago in the 1950s, and has since undergone numerous technical iterations and applications. In recent years, there has been a resurgence of AI technology and application, which can be observed from two perspectives. Firstly, there has been a rapid increase in the number of media reports and research on AI since 2014 (see figure 1), indicating that the public is paying more attention to this theme. Secondly, research conducted by institutions and media shows that organizations have a growing interest and are investing more capital in artificial intelligence. This trend is expected to continue<sup>1</sup>.

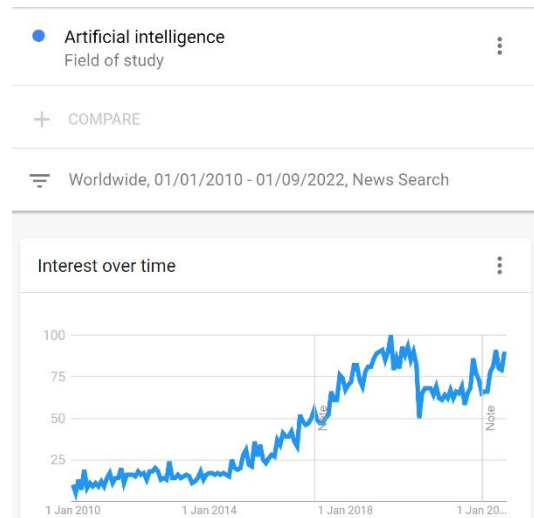


Figure 1: The number of media reports and research has increased rapidly since 2014

With increasing public attention and popularity of practical application of intelligent technology, governments and technology enterprises around the globe have drafted and published documents on ethical principles relating to artificial intelligence, and such efforts have ended with the publication of more than 100 documents<sup>2</sup>.

### A topic to be further explored

Our analysis of the corpora of documents of ethical principles on AI reveals an evident convergence in the commitments of diverse institutions to AI principles<sup>3</sup>. However, these documents are usually general guidelines and there is little reference to specific AI applications and scenarios, or the mechanism of implementation of norms or principles. Indeed, studies show that AI ethics is still a young field with few well-established issues and overviews<sup>4</sup>.

Firstly, most extant research, which employ mainly quantitative methods, considers either

<sup>1</sup> Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). *Notes from the AI frontier: applications and value of deep learning*. <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>

<sup>2</sup> Algorithm Watch. (2020). *AI ethics guidelines global inventory*. <https://inventory.algorithmwatch.org/>

<sup>3</sup> Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399. <https://doi.org/10.1038/s42256-019-0088-2>

<sup>4</sup> 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>

Müller, V. C. (2021). *Ethics of artificial intelligence and robotics*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/>

the value-principles or "application technologies and scenarios", but rarely takes both factors into account<sup>5</sup>. Secondly, some studies only explore an individual AI application (e.g. human recognition) and a single ethical value (e.g. privacy), instead of the broader types of AI applications and other ethical values/principles<sup>6</sup>. Thirdly, regardless of their quantitative or qualitative nature, the people's perception and attitude on AI of previous studies usually are wildly dependent on respondents' knowledge and perception of artificial intelligence, assume unformal understanding and thus lacking proper control on the knowledge conditions to their responses<sup>7</sup>. With the rapid pace of technological change and unique social background, it is difficult for researchers to measure the impact of respondents' knowledge and perception of artificial intelligence on the answers, and for this reason, some research findings will require further verification<sup>8</sup>. Finally, there are relatively few studies that focus on Hong Kong or Asia. With the rapid advancement and interest in AI application worldwide, research nested in the Hong Kong context is deemed to be warranted<sup>9</sup>.

### 1.3 Research objectives

To address the issues mentioned above, this research makes reference to a series of AI ethical principles and related domestic and overseas discussions, and embarks on an in-depth exploration, with the aim to:

1. Identify the ethical values that Hong Kong citizens are mostly concerned about in the applications of AI;
2. Understand the general views of the public on these ethical values and principles; and
3. Explore the attitude of Hong Kong citizens when faced with potential value conflicts in difference scenarios of AI application.

This study will consider both "ethical values" and "scenarios of AI application", focusing on the empirical descriptions and analysis of Hong Kong citizens' perceptions and opinions on

---

<sup>5</sup> Salesforce. (2018). *Artificial intelligence in Asia: trust, understanding and the opportunity to re-skill*.

<sup>6</sup> Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, 563(7729), 59-64. <https://doi.org/10.1038/s41586-018-0637-6>

<sup>7</sup> Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). *Notes from the AI frontier: applications and value of deep learning*. <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>

<sup>8</sup> Research of Salesforce pointed out that the more Hong Kong citizens understand, the less they are likely to trust AI applications. If this is true, then we should rethink the rationale for education and technology literacy.

<sup>9</sup>Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., Lyons, T., Etchemendy, J., Grosz, B., & Bauer, Z. (2018). *The AI index 2018 annual report*. Stanford University. [https://hai.stanford.edu/sites/default/files/2020-10/AI\\_Index\\_2018\\_Annual\\_Report.pdf](https://hai.stanford.edu/sites/default/files/2020-10/AI_Index_2018_Annual_Report.pdf) China emphasizes the machine aspect and relatively ignores humanities research, Hong Kong included.

the application of AI, and provide evidence-based preliminary policy recommendations.

## 1.4 Working definition of terms

### Artificial intelligence

We adopt the working definition<sup>10</sup> of AI provided by the OECD, which defines AI as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy"<sup>11</sup>. The AI applications and scenarios analyzed in this study have specific goals.

### Ethical values, principles, and norms

Ethical values are moral beliefs that provide motivation, be they driven by needs or motivated by incentives. When these ethical values follow specific rules of judgment and are applied in real-life situations, they become ethical principles. Ethical norms are ethical principles practiced under specific mechanisms and power structures. A value can be realized by a variety of rules of judgment.

For example, the value of "robustness" is a universal motivation based on human need for security, minimizing errors, and the tendency of risk aversion. In the field of engineering or product safety, robustness requirements become principles when they meet specific standards or attain a satisfactory quality level. Depending on the scenario, the type of technology application, or local cultural differences, this robustness principle may not always be universally applicable, in other words, it is not a one-size-fits-all standard.

In nuclear energy applications, the European region expects relevant institutions to adopt the precautionary principle to deal with nuclear risks, while the same principle does not apply to Teflon non-stick pots, even though we currently have no way of knowing how high the long-term carcinogenic risk is in actual household usage.

Finally, this mechanism becomes the norm when institutions (usually, but not necessarily, governments) demand uniform safety standard requirements by legal or market forces and subsidies or penalties are given out to producers. Therefore, norms are most closely related

---

<sup>10</sup> A working definition differs from a definition. The former is an explanation of a term or concept that provides operational interpretation (interpretations). A concept can have different interpretations or working definitions depending on the purpose of the study. A definition, on the other hand, allows only one interpretation.

<sup>11</sup> OECD. (2022). *Recommendation of the Council on artificial intelligence* (OECD/LEGAL/0449). <https://legalinstruments.oecd.org/api/print?ids=648&lang=en>

to governance because governance cannot work against norms but is executed under the guidance of the latter.

## Governance

The term governance has various meanings, but what they all have in common is that governance is not uniquely governmental. Structures, corporations, and individuals outside the government often play important roles in governance<sup>12</sup>. As a political scientist has noted, “[g]overnance is a very old concept, and an even older reality... Governance is not a constant, but rather tends to change as needs and values change”<sup>13</sup>.

AI governance involves the identification of problems and controversies that arise from various areas of AI applications, as well as the underlying needs and values. They are then conceptualized into specific regulatory and policy issues before finding a suitable and feasible mechanism or role to address these problems<sup>14</sup>.

For instance, in large-scale public construction projects like undersea tunnels, the government may adopt public-private partnership and build-operate-transfer models to reduce the huge initial costs and risks of undertaking the project alone. In this case, a joint venture of government and private enterprises forms a suitable and feasible mechanism or role combination that changes in accordance with factors such as project needs (e.g. expenditure and risk), government financial capacity, or other values (e.g. environmental protection and EIA reporting). Yet, there is no constant formula for simple replication.

## 1.5 Literature review

### 2.2.1 Literature research (I): Global perception of artificial intelligence technology

The development of artificial intelligence (AI) has been a game changer for many industries. The emergence and advancement of algorithm-based technologies has led to the

---

<sup>12</sup> “Governance can be used as a specific term to describe changes in the nature and role of the state following the public sector reforms of the 1980s and 1990s. Typically, these reforms are said to have led to a shift from a hierarchical bureaucracy towards a greater use of markets, quasi-markets, and networks, especially in the delivery of public services.” “By analogy, governance also can be used to describe any pattern of rule that arises either when the state is dependent upon others or when the state plays little or no role.” In Bevir, M. (2009). *Key concepts in governance*. SAGE.

<sup>13</sup> Pierre, J., & Peters, B. G. (2005). Governance: A Garbage Can Perspective. In *Governing complex societies: trajectories and scenarios* (pp. 49-63). Palgrave Macmillan UK.

[https://doi.org/10.1057/9780230512641\\_3](https://doi.org/10.1057/9780230512641_3)

<sup>14</sup> Ibid.

rise of new industries and unveiled groundbreaking potential for the society as a whole<sup>15</sup>. Due to its transformative capabilities, AI technology has an immense impact on different aspects of modern society, including education, transportation, security, the commercial sector, and healthcare. In addition, as these technologies continue to grow and develop, more benefits and possibilities are yet to be unveiled in future<sup>16</sup>.

However, there are growing concerns that the expanding power of algorithm-based technologies and the pervasiveness of AI-related innovations may pose certain risks to the balanced development of society. With loose regulations in the field and a potential conflict of interest among the government, developers, and the commercial sector, the technology can potentially bring negative consequences to user privacy and data integrity, leading to devastating outcomes, such as greater inequality and discrimination.

The society can be seen as an amalgamation of different stakeholders in the development and deployment of AI technologies. Considering both the potential benefits and risks of AI's expansion, it is important to understand the societal perceptions of this new technology<sup>17</sup>.

### **Perception of AI in the United States**

The United States is one of the leading countries in the development of AI-related technologies. While current regulations and the vigilance of civil society restrict the uncontrolled deployment of algorithm-based innovations, there is a tendency for commercial companies to deploy AI more extensively, leading to increasing debates over the technology's future in the United States<sup>18</sup>.

Societal attitudes towards AI technology reveal that the general public supports its development, highlighting its benefits at the current level of technology deployment. A study has found that 41% of the respondents held the belief that technology is beneficial to society, and they showed a positive attitude towards its different applications. However, some other studies on public perception of AI in the United States have identified opposing

---

<sup>15</sup> Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, 2(1), 10-12. <https://doi.org/10.1038/s42256-019-0136-y>

<sup>16</sup> Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature (London)*, 568(7751), 163-165. <https://doi.org/10.1038/d41586-019-01099-5>

<sup>17</sup> Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI*.

<sup>18</sup> Surya, L. (2017). AI in information technology and its future in the United States. *International Journal of Creative Research Thoughts (IJCRT)*, ISSN, 2320-2882.

trends regarding AI advancement, although respondents did not perceive it as a dominant issue for the upcoming ten years<sup>19</sup>.

Recently, new reports from the United States have raised concerns across the population about the proliferation of AI technology. The latest survey conducted by Pew Research Center indicates growing concerns over the increased use of AI in daily life, with the feeling of concern prevailing over excitement across 37% of the American respondents. In addition, the majority of survey participants are equally concerned and excited about the extensive use of AI in daily life<sup>20</sup>.

**Demographic and preference analysis has identified key parameters, such as gender, age, education, religious affiliation, and income as important predictors of AI perception among the US population<sup>21</sup>. Certain social groups are more predisposed to support the development of AI**, specifically male respondents from younger age groups with higher levels of education, non-religious affiliation, and higher income backgrounds. The demographic results also show that there is a substantial difference in the perception of AI technologies among US citizens with different political affiliations: Democrats tend to voice higher support for algorithm-based innovations than Republicans.

The analysis also shows that US citizens perceive different AI-based technologies in varying ways. For example, most respondents view the use of AI in facial recognition by police forces positively, with only 27% of them voicing concerns over the technology. Similarly, the algorithm-based fact-checking programmes used by social media companies have a high acceptance rate among the respondents. However, American citizens display a greater mistrust of autonomous vehicles, with more than 44% of the citizens expressing concern over the technology. This mistrust is due to the vulnerability of driverless cars to cyber-attacks and the possible negative impact on the job market. The majority of respondents also do not feel safe delegating immense authority to algorithm-based mechanisms in decision-making processes<sup>22</sup>.

---

<sup>19</sup> Northeastern University and Gallup. (2018). *Optimism and anxiety: views on the impact of artificial intelligence and higher education's response*. Gallup.

<https://www.northeastern.edu/gallup/pdf/OptimismAnxietyNortheasternGallup.pdf>

<sup>20</sup> Pew Research Center. (2022). *AI and human enhancement: Americans' openness is tempered by a range of concerns*. Pew Research Center. [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS\\_2022.03.17\\_AI-HE\\_REPORT.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS_2022.03.17_AI-HE_REPORT.pdf)

<sup>21</sup> Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. University of Oxford. [https://isps.yale.edu/sites/default/files/files/Zhang\\_us\\_public\\_opinion\\_report\\_jan\\_2019.pdf](https://isps.yale.edu/sites/default/files/files/Zhang_us_public_opinion_report_jan_2019.pdf)

<sup>22</sup> Pew Research Center. (2022). *AI and human enhancement: Americans' openness is tempered by a range of concerns*. Pew Research Center. [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS\\_2022.03.17\\_AI-HE\\_REPORT.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS_2022.03.17_AI-HE_REPORT.pdf)



Studies focusing on public perception of technology in the United States highlight differences in attitudes towards AI across the American public. While a large portion of the population supports the development of technology, certain social groups reveal a lower degree of support to the further development and deployment of AI, such as pre-Millennials, females, as well as persons with lower family income, lower educational attainment, and no computer science and programming experience. Moreover, **the acceptance rate of new technology fluctuates significantly across different types of AI applications**, with highly automated and decision-making processes receiving lower support from the American society<sup>23</sup>.

### Perception of AI across Europe

In Europe, the development of new AI technologies and their growing power and presence across different spheres of everyday life has changed the balance between stakeholders. The widespread use of AI technology across the private sector has become a major point of concern for civil society as weak regulations might spark waves of data leakages or systematic technological abuses by commercial companies. Thus, the current stage of AI development in Europe focuses on legal regulation of technology, addressing the challenges member countries might face due to the uncontrolled spread of algorithm-based technologies<sup>24</sup>.

European societies attach great importance to the protection of personal data and recent debates on the security concerns of AI have negatively affected the image of the technology. Therefore, the deployment of AI technology, especially in public spaces, has been significantly restricted. At the same time, public perception of AI across European countries is similar to that in the United States and divided opinions over the future of the technology exist among different social groups<sup>25</sup>. Across the EU community, **the perception of AI technologies differs significantly depending on factors such as gender, income, education, and professional knowledge of AI**. Gender plays an important role in the perception of AI, with male respondents voicing much higher support for the development of the technology<sup>26</sup>. Some recent reports have identified an even greater gap across gender,

---

<sup>23</sup> Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. University of Oxford. [https://isps.yale.edu/sites/default/files/files/Zhang\\_us\\_public\\_opinion\\_report\\_jan\\_2019.pdf](https://isps.yale.edu/sites/default/files/files/Zhang_us_public_opinion_report_jan_2019.pdf)

<sup>24</sup> MacCarthy, M., & Propp, K. (2021, May 4). *Machines learn that Brussels writes the rules: The EU's new AI regulation*. Brookings. <https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/>

<sup>25</sup> Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. University of Oxford. [https://isps.yale.edu/sites/default/files/files/Zhang\\_us\\_public\\_opinion\\_report\\_jan\\_2019.pdf](https://isps.yale.edu/sites/default/files/files/Zhang_us_public_opinion_report_jan_2019.pdf)

<sup>26</sup> European Commission. (2017). *Special Eurobarometer 460: attitudes towards the impact of digitisation and automation on daily life*. European Commission.

with male respondents showing twice as much support for AI than their female counterparts<sup>27</sup>. Highly educated people support the development of technology much more often, while lower-educated ones stress the possibility of job losses caused by the integration of new technology into industries. High-income individuals demonstrate greater acceptance, while low-income respondents express greater concerns over the development of AI<sup>26</sup>. In addition, respondents holding a degree in IT or having attended educational courses in the past express much higher trust in AI technology<sup>27</sup>.

Another particular concern in the perception of AI technologies across European countries is the attitude towards AI-based applications that collect biometric data. Respondents from European countries reveal a more negative attitude towards facial recognition technology in comparison to other countries studied in the research. There is fear that such AI-based application will bring new risks rather than benefits to everyday life<sup>28</sup>. The study results support a more cautious attitude towards AI across European countries, prioritizing public attention to anticipated risks over the expected benefits of technology.

## 2.2.2 Literature research (II): Global perceptions of AI technology and the ethical values of AI

Research indicates that ethical concerns vary across countries. For example, fairness and transparency tend to be given a high priority by the British, Americans, and Europeans, whereas safety is emphasized more by the Chinese<sup>29</sup> <sup>30</sup>. In the field of artificial intelligence, Europe and the United States have invested heavily in humanities and social science research. While China is seen to be catching up in AI studies in social science research, it is still lagging behind in the field of humanities<sup>31</sup>. Since 2019, researchers have found that institutions all over the world have made a significant effort to commit themselves to

---

<sup>27</sup> Sartori, L., & Bocca, G. (2022). Minding the gap(s): public perceptions of AI and socio-technical imaginaries. *AI & society*. <https://doi.org/10.1007/s00146-022-01422-1>

<sup>28</sup> Kostka, G., Steinacker, L., & Meckel, M. (2021). Between security and convenience: facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. *Public Understanding of Science*, 30(6), 671-690. <https://doi.org/10.1177/09636625211001555>

<sup>29</sup> Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., & Niebles, J. C. (2019). *The AI index 2019 annual report*. Stanford University. [https://hai.stanford.edu/sites/default/files/ai\\_index\\_2019\\_report.pdf](https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf)

<sup>30</sup> Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, 563(7729), 59-64. <https://doi.org/10.1038/s41586-018-0637-6>

<sup>31</sup> Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., Lyons, T., Etchemendy, J., Grosz, B., & Bauer, Z. (2018). *The AI index 2018 annual report*. Stanford University. [https://hai.stanford.edu/sites/default/files/2020-10/AI\\_Index\\_2018\\_Annual\\_Report.pdf](https://hai.stanford.edu/sites/default/files/2020-10/AI_Index_2018_Annual_Report.pdf)

upholding ethical values in AI, with a certain degree of consensus around the values they pledge to support<sup>32</sup>. These values include:

- Transparency
- Justice and fairness
- Non-maleficence
- Responsibility
- Privacy
- Beneficence
- Freedom and autonomy
- Trust

However, some researchers argue that a list of widely recognized moral values is insufficient as it may not address everything in a developing issue. They suggest that these values represent only a superficial consensus with little practical significance, and therefore further elaboration and implementation are considered necessary<sup>33</sup>.

### 2.2.3 Literature research (III): A preliminary summary of research directions

Based on the two groups of literature above, we have come up with four research directions:

1. Investigating the relationship of different demographic background of Hong Kong citizens, including the impact of gender, age, and education level, with their perceptions of artificial intelligence, using questionnaires.
2. Examining potential issues arising from different types of AI applications (such as autonomous vehicles) in various contexts, and observing how public perception changes in response to variations in application types and scenarios.
3. Understanding how citizens comprehend the range of ethical values and principles raised in global discussions on AI governance.

---

<sup>32</sup> Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI*.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399. <https://doi.org/10.1038/s42256-019-0088-2>

Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking artificial intelligence principles. *arXiv preprint arXiv:1812.04814*.

<sup>33</sup> Jobin, A., Man, K., Damasio, A., Kaissis, G., Braren, R., Stoyanovich, J., Van Bavel, J. J., West, T. V., Mittelstadt, B., Eshraghian, J., Costa-jussà, M. R., Tzachor, A., Jamjoom, A. A. B., Taddeo, M., Sinibaldi, E., Hu, Y., & Luengo-Oroz, M. (2021). AI reflections in 2020. *Nature Machine Intelligence*, 3(1), 2-8. <https://doi.org/10.1038/s42256-020-00281-z>

4. Exploring the roles of AI governance, including those of the government and private enterprises.

Research directions	Semi-structured interviews	Questionnaires
Direction (1)		✓
Direction (2)		✓
Direction (3)	✓	✓
Direction (4)	✓	✓

Before distributing the questionnaire, a semi-structured interview was conducted to collect preliminary comments on the questionnaire and to assist us in interpreting the results of the questionnaire through more in-depth interviews.

## 1.6 Research methods

### 2.2.4 Semi-structured interviews

We initially aimed to invite individuals from Hong Kong's innovation and technology industry, business sector with management experience, and the academia to participate in focus group studies by the snowball sampling method. Due to the coronavirus pandemic, the plan was modified to adopt individual virtual and face-to-face interviews, taking place between May and November 2021. We have conducted a total of 22 individual interviews. The interview was semi-structured, with interviewees answering questions in the interview protocol (see Appendix A) and the interviewers asking additional questions based on the specific content of the conversation. The outline was structured around two themes, "Areas of Concern on Smart City and Automation" and "Corresponding Data Governance Framework". At the end of the interview, interviewees were given a draft of the questionnaire and asked for their opinions, so as to reduce the interviewers' intervention and bias on the interview content while keeping the outline of the conversation open-ended. In other words, the content of the interview might go beyond what was covered by our questionnaire.

### 2.2.5 Questionnaires

In March 2022, the research team commissioned the Centre for Communication and Public Opinion Surveys (CCPOS) to conduct a telephone survey on ethical values and principles in AI governance from late April to early May, with a random sample size of 500 (see

Appendix B). This telephone questionnaire survey explored **citizens' perceptions of different types and scenarios of AI applications and related ethical values, with specific behavioural experiences excluded**. As AI applications could be unfamiliar to some people of Hong Kong, the questionnaire was designed not to require respondents' knowledge of AI applications (e.g., "Do you know what AI applications exist in daily life?").

Problems were broadly divided into seven categories (see Appendix C), the first two of which did not provide a description of intelligent applications and scenarios:

1. Decontextualized value ranking
2. Decontextualized value trade-off

Although more than five sets of ethical values or principles are found in existing literature, we have only selected five sets with some wording modifications (e.g., the more abstract "fairness and justice" was changed to the more specific "unbiased," and "freedom and autonomy" was changed to "individual freedom") based on reasons like the clarity of these concepts, the ranking of the importance of these values or principles in previous research, the number of times they were mentioned in interviews, and the limitations of research resources.

The next four categories encompassed four scenarios:

3. Whether specific AI applications should be adopted in health codes, fraud detection systems, and driverless autonomous vehicles
4. Acceptance of specific AI applications in robotic dog surveillance
5. Situational value trade-off in specific intelligent application technologies and scenarios
6. Whose opinion should be prioritized in specific intelligent application technologies and scenarios

The scenarios were modified from existing literature and actual cases that happened elsewhere to maintain some authenticity. The adoption of smart lampposts was not included in the scenarios to avoid over-amplifying its impact<sup>34</sup>.

The first two types of questions only addressed the issue of "ethical values", while both "specific application technologies and scenarios" and "ethical values" were areas covered in

---

<sup>34</sup> Hong Kong experienced rare civil unrest in 2019. During the unrest, some lampposts were torn down because they were suspected to have surveillance and facial recognition functions. The Guardian. (2019). Hong Kong riot police beat protesters at anti-surveillance rally. *The Guardian*. <https://www.theguardian.com/world/2019/aug/24/hong-kong-fresh-rallies-as-protesters-target-airport-transport>

the last four.

The last category collected demographic data for analysis, with the aim to offer demographic analysis and to compare results with corresponding overseas research.

Interpreting the results of the questionnaire often involves three basic questions:

1. What are the preferences/principles reflected by the respondents' judgments on value issues?
2. Do the respondents offer consistent responses to the value questions?
3. If not, under what circumstances does this consistency begin to change?

## Semi-structured interviews

The purpose of the interviews is to gain an initial understanding of AI ethics and governance in the context of Hong Kong, and to capture the unique characteristics of local discussions. From May to November 2021, via snowball sampling method, individuals from the local technology industry, business sector with relevant management experience, and the academia were invited to participate. Of the 22 interviewees, 11 worked in information and communications technology (ICT) or innovation and technology (I&T), 7 had legal or relevant management experience, and 4 were from the academia. Many respondents working in the ICT industry also had management experience. The interviewees were free to respond to the protocol questions. During the associative thinking questions at the beginning of the interview, interviewees were encouraged to think and answer freely according to their perceptions and impressions, and additional guidance was only provided when clarification of a question was requested. In the middle of the interview, we extracted concepts offered by individual respondents for follow-up questions, and paraphrased, summarized, or reorganized the concepts in the conversation to ensure accuracy. In the final part of the interview, respondents were invited to answer and comment on the English draft of the telephone interview questionnaire.

### 2.1 Smart cities and AI applications (that arouse areas of concern)

#### 2.2.6 Background of the interview

In the five years leading up to the interview, organizations at all levels around the world, including government agencies, had successively published AI strategies, ethics and governance frameworks, and drafts of legislation. The Hong Kong Special Administrative Region Government published the first Smart City Blueprint in 2017 and the follow-up 2.0 version in 2020, aiming to strengthen the infrastructure and talent training of AI technology to unleash its potential. The blueprint serves as a policy directive and a showcase of the projects undertaken by the government.

In this report, we explore the issues related to technology ethics, governance, and legislation, instead of focusing solely on technical and economic perspectives. We aim to see if people's perceptions of these issues align with or diverge from global standards. Especially after the unique social atmosphere in Hong Kong in 2019, lessons from the destruction of the experimental smart lamppost project have served as a warning for us not to underestimate the impact and the controversy that may arise from public perception and

technology ethics that are beyond the technical and economic perspectives.

To avoid overemphasizing individual events, we did not mention the smart lamppost incident during the interview or telephone questionnaire sessions. Furthermore, we conducted the interviews nearly two years after the incident, allowing ample time for the situation to settle and the sentiment to cool down. Nevertheless, during the interviews, some respondents mentioned that the incident is still worth studying in the development of local intelligent applications. Some even pointed out that the reason why some AI technologies have not been successfully implemented is not due to technical defects, but rather due to the distrust of technology users. One respondent stated,

*"It is not about technology, it is just that citizens do not trust (the government) ... These are not technical issues at all."*

Others noted that this **distrust stems from a lack of thorough explanation to and communication with the users of the technology before the implementation of smart applications**. The general public often do not understand the functions and implementation details of the technology. For instance, one respondent said,

*"Obviously, no public consultation was conducted before the case of smart lamppost... So communication with stakeholders is needed, and a bottom-up approach is needed."*

Finally, another respondent highlighted **the importance of educating the public and promoting transparency in technology projects**.

*"In the early years, everyone saw that the lampposts were torn down or damaged, and it is not a matter of technology, it is about how to educate the people and promote the technology. The lesser information is available to the public, the more psychologically people feel that there is a conspiracy behind it. As an ordinary citizen, I really do not think that the smart lamppost project was transparent enough to let you know the makeup of the lamppost technology, and how information was collected and used. The government only let you know that the smart lamppost system could do a lot and what could be monitored. People will be even more frightened when they hear the term 'monitor'."*

### 2.2.7 Demand-driven smart cities

Smart cities are characterized by the use of digital technologies, knowledge, and smart



assets to quickly respond to citizens' needs, improve city services, and make cities more livable. One of the biggest challenges for any city's stakeholders is to identify a development direction and create measurable benchmarks to lead an efficient, high-performing city while boosting economic, social, and cultural vitality. It all starts with how the idea of "smart" is conceptualized and operationalized.

In its latest study titled "Smart Cities: Key Technologies, Environmental Impact and Market Forecasts 2022-2026", Juniper Research has formulated its smart city ranking based on an extensive study of cities around the world, with the top five smart cities being Shanghai, Seoul, Barcelona, Beijing, and New York<sup>35</sup>. The ranking of the world's 50 cities is based on an assessment of many different aspects, covering transportation and infrastructure, energy and lighting, urban management and technology, and urban connectivity. The research also focuses on the rapid deployment of data management platforms, efficient and digital utility management, and public services. Many Asian cities have taken a similar approach, assessing basic infrastructures, public services, and socio-economic conditions. An example is Hong Kong's smart city blueprint, which has six development areas: "Smart Mobility", "Smart Living", "Smart Environment", "Smart Citizen", "Smart Government", and "Smart Economy". The Juniper Research Ranking is more closely related to three of the six categories, namely "Smart Mobility", "Smart Environment", and "Smart Government", while less related to the remaining three. This difference indicates that the concept of a smart city is not as solid as one might think and may need further consolidation for universality or to be accepted and understood as a concept that is city-specific.

In the interviews, respondents were asked to think associatively about "the application of artificial intelligence in smart cities" and briefly express their understanding of "artificial intelligence" and "smart city" in Hong Kong. Some respondents immediately mentioned the six major areas of the government-led smart city blueprint, such as "Smart Mobility". Others thought of the smart applications developed by I&T companies to improve living conditions. Others believed that smart cities and AI are just empty rhetoric or buzzwords that lack substantive meaning. Although the responses varied, the respondents generally agreed that these concepts have to be based on real substantive needs.

*"All technical applications require a certain level of demand - without demand, even if the technology is good, it does not make sense."*

---

<sup>35</sup> Moar, J., & Bainbridge, M. (2022). *Smart cities: key technologies, environmental impact and market forecasts 2022-2026*. J. Research. <https://www.juniperresearch.com/researchstore/healthcare-government/smart-cities-research-report>

However, survey participants had varying interpretations of the terms "demands" or "needs". Some respondents viewed it as an international competition between cities, while others saw it as a need to improve work efficiency and quality of life. Those who believed that smart cities and artificial intelligence lack substance pointed out that Hong Kong lacks substantive research or consultation on these needs, and that the acceptance level towards technology is not satisfactory. The low acceptance level towards technology expressed by some interviewees echoes the findings of IMD and the Singapore University of Technology and Design (SUTD) on the Smart City Index. The index study examines the economic and technological aspects of smart cities on the one hand, and the quality of life, environment, and inclusiveness of smart cities on the other. The agency responsible for the study surveyed citizens of 118 cities in July 2021 and found that 45.8% of the Hong Kong respondents were comfortable with the use of facial recognition technology to reduce the crime rate. Among them, 53.2% also believed that the availability of online information increased their trust in authorities. The proportion of respondents using non-cash daily payment transactions (transaction percentage) was 69.4%, and the percentage of respondents willing to give up personal data to improve traffic congestion was 59.9%. All these values, except the last one, are below the average level of acceptance in other cities<sup>36</sup>.

Interviewees with legal or related management experience tended to believe that the key to a smart city lies in the "soft" component behind the "hardware" technology, that is, development of smart cities from the perspective of urban management. City managers believed that there are several key tools for effective smart city development, and cities can use different ways to achieve smart city outcomes. Much research, especially related to AI, focuses on talent preparation, people-centred needs, innovation ecosystems, and smart policy. Based on the above understanding of demand, respondents thought it advisable to start with the popularization of talent preparation and people-oriented needs, in response to the lack of an innovation ecosystem and public understanding. They had the opinion that the government seems to lack the experience with technology-related policies, from promotion to implementation. For example, they doubted the success of the implementation of some previous policies introduced in Hong Kong, such as the automatic number plate recognition system for vehicles (in electronic road pricing scheme), smart litter containers, and smart lampposts. A mismatch of demand and resources was also seen in the application of technology.

---

<sup>36</sup> Institute for Management Development and Singapore University for Technology and Design. (2021). *Smart city index 2021*. Institute for Management Development. [https://imd.widen.net/s/x25tgd8drv/smart\\_city\\_index\\_2021](https://imd.widen.net/s/x25tgd8drv/smart_city_index_2021)

## 2.2.8 Understanding of intelligent applications

### 2.1.3.1 Common misconceptions... AI is inevitably related to big data

Some of the respondents with an ICT or I&T background tried to clarify some common misconceptions about AI. One of these misconceptions is that **AI technology inevitably relies on big data**. In fact, developers of intelligent technologies have different methods to work around this limitation without necessarily relying on large amounts of real data. The functions and impacts of such intelligent applications are not necessarily less than those that rely on large amounts of data. Another misconception is that **institutions have unlimited processing capacity for big data, reflecting a failure to consider limitations such as storage and computing costs**. The final misconception is that any data has value for analysis. Despite this, it is generally agreed that intelligent technologies today require data more than they did in the past.

### 2.1.3.2 "Seeing harms before the benefits": Attention asymmetry to functions and technical details

Some respondents pointed out that good technology is so entrenched in our daily life that users actually do not recognize that the benefits generated are actually a product of artificial intelligence. Representative examples cited in the interviews include Optical Character Recognition (OCR) and Google Maps. Both technologies have been widely used for years, and many regular users are not concerned whether their working mechanisms involve AI technology. Such examples seem to illustrate a **phenomenon of asymmetry**: When a user is familiar, function-wise, with a technology, the part of the technology that involves AI technology will be neglected; on the contrary, when faced with new technologies, if the user is not familiar with their functions or how they can facilitate or meet his/her needs, the technical details involved in the AI technologies would be put more under scrutiny.

As a result, it seems easy for people to generate a negative association intuitively towards AI technology. Some respondents frankly said that "they do not see the benefits, but the (potential) harms first". Such phenomena or opinions observed from the examples do not necessarily establish a causal relationship, and when we try to promote or educate the public about AI technology (including highlighting the nature of intelligent technology), it may not automatically change their attitude towards the function of technology (e.g., induce them to doubt the function of technology). Yet, many respondents still believed

**good technology literacy can reduce the public's skepticism and unnecessary backlash towards technology**; they thus believed Hong Kong should carry out relevant work as soon as possible, which would be helpful in promoting I&T development.

### 2.1.3.3 Social tendencies to focus on vague advantages but specific disadvantages

**When discussing the pros and cons of AI technology, respondents generally provided fewer and less detailed descriptions of the advantages as compared to the disadvantages.** Some of the advantages commonly mentioned include improving quality of life, increasing efficiency, providing personalized experiences for users, and reducing dependence on humans. They also believed that AI technology can result in fewer conflicts caused by human error and communication misunderstandings. However, it is important to note that these advantages are not universally applicable. For example, respondents thought that younger customers may prefer self-service ordering machines rather than relying on store staff, while older customers may prefer traditional shop assistants taking orders for them manually. Similarly, chatbots for customer service may satisfy younger customers' tendency to reduce dependence on human service and thus can avoid direct interpersonal conflict, but elderlies may not feel the same way.

Compared to the advantages, the shortcomings of AI technology mentioned are more specific and may pose serious challenges. Respondents pointed out some controversial examples abroad, such as Cambridge Analytica's improper manipulation of public opinion during the election campaign, to illustrate the adverse effects of personalization. They thought that even if the Clearview face recognition system is assumed to have a 98% accuracy rate, the remaining 2% of false judgments can still cause detrimental consequences when used in law enforcement. Others believed that social media recommendation systems might create echo chambers that reinforce users' biases against specific groups. Some respondents also mentioned that local applications such as smart lampposts or the early LeaveHomeSafe app were not well-received due to their new and sophisticated nature, lack of easy-to-understand explanations, and time required for public adaptation.

In addition, there were concerns that AI technology may exacerbate the problem of responsibility evasion. For example, the accountability framework behind technology may be unclear, or the chain of responsibility may become too long after the introduction of AI technology. Additionally, it was feared that some people may shirk responsibility for technology immaturity. Respondents also pointed out that AI applications may lead to the

loss of tacit knowledge involved in service-intensive work, such as doctors and nurses visiting wards in person to record patients' conditions and progress or to train new staff. This loss of knowledge may be irreversible. Local small and medium enterprises (SMEs) that rely heavily on interpersonal communication and experience may face similar challenges.

In summary, respondents were generally more specific when describing the disadvantages of AI technology, while being less detailed when discussing the advantages. This captures the general view of AI in the society, bearing in mind that some of the respondents interviewed already possessed a deeper understanding of intelligent technology than the general population.

## 2.1.4 What people care about and what transparency means

### 2.1.4.1 Local responses that echo the global concern

During the interviews, the responses that echo the ethical values of global concern identified in the existing literature include personal privacy, transparency, safety, and fairness.

### 2.1.4.2 What the user/consumer thinks

Some respondents attempted to analyze how general users or customers perceive AI technology in terms of transparency. Some believed that users often feel helpless or even disgusted when facing AI technology. Three possible factors might lead to these emotions:

1. **Asymmetrical data ownership:** Institutions collect and process more information or data than individuals and individuals view this as an asymmetrical relationship;
2. **Absence of options:** Users/consumers of AI technology often do not understand what benefits new technologies bring to them, and feel that they have no other options available to them; and
3. **Feeling of being compelled:** Even if they agree with the benefits of technology and the necessity of its use in daily life, they may feel that they cannot control it but are controlled by it instead.

Regarding how to make users or customers feel in control, some respondents believed that users/customers of AI technology deserved to be paid for the data they share as a return. They were in the opinion that organizations should handle information or data with control

and should avoid abuse, such as not over-collecting or over-retaining it under the guidance for good privacy protection practice. The benefits of AI technology over older practices should also be clearly explained to users/customers and there should be options to maintain old practices (or transitional options) for those who are not ready to embrace the new technology yet.

#### 2.1.4.3 Other understandings of transparency

Some respondents viewed transparency as giving technical experts access to important data retained by organizations for review (mandated by policy) to clarify the relationship between authority and responsibility and to give suggestions for improvement. Others commented that large institutions, including the corporates and the government, may use privacy protection as an excuse to monopolize data usage for their interests. The monopolization makes it hard for small I&T start-ups to develop and compete. If the government can proactively share its data and require large enterprises to share some of their data, that will help to break the monopoly or reduce asymmetry of data owned by large enterprises. This, in turn, is beneficial to consumers and SMEs. Another respondent pointed to the example of healthcare in Hong Kong, where data sharing is mostly unidirectional from public to private practices. Data collected from private organizations is rarely shared with third parties.

#### 2.1.5 Summary: Areas of concern

To summarize the opinions of the respondents, the biggest challenge facing smart cities and new AI technology-related projects is not the lack of hardware infrastructure, but the lack of effective communication with stakeholders during deployment or implementation. Communication management should focus on understanding the real needs of the AI technology users and responding accordingly, rather than simply providing the advanced and ambitious technologies from a supplier perspective.

Moreover, communication management should also take into account the importance of technology education for recipients, which involves clarifying technical and functional misunderstandings and promoting awareness of intelligent technology. Improving technology literacy will reduce skepticism and instability, and minimize the obstacles faced by new technology projects.

Currently, society tends to view the advantages of intelligent technology rather vaguely, while the focuses on its disadvantages are more specific, resulting in an unbalanced discussion between the two. Differences in understanding of intelligent applications show

that the benefits or advantages of the technology are not effectively communicated across, while the disadvantages or drawbacks are not adequately addressed. Respondents thus felt that special attention to striking a balance in communication management should be the future focus for AI technology providers and managers.

The respondents' feedbacks also echo ethical concerns expressed around the world, with transparency being a particularly significant and widely discussed issue. This may be due to the fact that respondents in this study had more understanding of AI technology than the general public and tended to think more concretely about its implications. Transparency not only refers to the understanding of the technology, but also encompasses concerns around other values such as privacy, which may arise from a lack of understanding.

## 2.2 Appropriate data governance framework

### 2.2.1 Interpretation and education: Differences in understanding and expectations of experts

**Data governance** involves explaining and providing educational strategies for smart technologies. Respondents shared different views, which are not necessarily in conflict with each other. Some felt that developers of AI technologies are aware of the potential disadvantages of their technology but may not be able to explain them in plain language. Another group believed that technology developers (e.g., I&T enterprises) have sufficient internal expertise to deal with ethical and governance issues, but the general public does not understand them. For example, when most users do not care to read privacy statements, it does not help if companies try their best to provide a reasonable explanation, thus assigning all the blame to the companies is not fair.

On the other hand, some respondents believed that practical applications in society are too complicated for researchers to anticipate all potential harms, let alone prevent them. Therefore, they thought that it is preferable and more effective for the government to provide specific frameworks and guidelines. Others suggested that a mechanism should be set up for third-party/external experts to validate and explain intelligent systems when needed (e.g., during an error incident), to accumulate experiences from individual cases, and to educate the public. Although respondents had different interpretations of AI technology and educational strategies, one commonality is found among their responses:

**The task of education and interpretation should be carried out by experts or authorities, and the general public who lack technical knowledge can only be the passive recipients of education.**

A smaller number of respondents offered a **slightly different view**: While they also felt that expert opinions are crucial, **these experts are framed as consensus builders or knowledge creators for governance**. One respondent pointed out the extreme information asymmetry between the general public on one hand, and enterprises and institutions on the other, and the former's lack of discursive power under the asymmetry. Even if the general public has a fragmented understanding of the topic, perhaps with strong feelings and thoughts, "*the challenge of governance today is that there is too much information*", as one interviewee put it. Such fragmented understanding of the public poses obstacles in establishing an effective conceptual framework to address governance issues directly. **A more practical approach suggested by the respondents is to build more consensus through the involvement of both the general public and experts, working together interactively for greater discursive power, instead of passive and unilateral dissemination of knowledge**. This approach allows parties to learn from each other.

Another interviewee suggested a more specific strategy: When promoting AI technology, it is better to let some members of the ordinary public understand and explore the benefits of the technology functionally by themselves, rather than having developers and experts explain the application of the technology directly. Letting users share their experiences with other users can result in good product feedback and communication. The process is not simply conveying the right information by experts, but the building up of confidence in the technology by the general public and the experts in an effective and interactive way.

*"It is common for laymen to be hesitant when they are overwhelmed with technical details and explanations, which turn out to trigger worries. Can we do things the other way round and let the users share their user experience with another user? It is better than a developer speaking by himself."*

At the same time, many respondents believed that third-party professionals, including universities and public institutions (such as statutory bodies that are not directly under the government), can play a crucial role in explaining and educating the public about AI applications, especially when it comes to government applications. Such opinions are compatible with the idea of building an interactive and trusting communication framework.



*“... Overall there is a principle, for every important data collection process, there are checks and balances... If the law can be written to a certain extent that enables data collection while allowing third-party expert verification, I think that will be safe for everyone.”*

*“I think that it will certainly be better for things to be done by the government, or even by universities. The transparency of private institutions is limited. People always mention WhatsApp and Facebook these days for collecting your data and feeding you advertisements in return...”*

*“Even if you make things transparent, people still believe in conspiracy... If you want people to believe you, it's actually important to be able to participate, inviting parties, be they children, universities, or professional parties. When people can take part in it, it will be more transparent than doing everything on one's own.”*

The results of this section are different from the survey results, and the reasons for the differences will be compared in detail in the “Discussion” section.

## 2.2.2 Hong Kong's positioning and characteristics, and the role of the government

The characteristics of Hong Kong's market for AI have affected respondents' views on an appropriate governance framework. Some respondents believed that the market demand in Hong Kong is too small. When local I&T enterprises develop their products, they usually target the global market, not just the market of Hong Kong. Businesses tend to see Hong Kong as a small-scale testing ground or a relay point for mainland companies looking for the international market:

*“The more regulations there are, the more boundaries there will be to limit developers' innovation. (Take fintech as an example)... Hong Kong's regulatory mechanism is comprehensive, which makes development relatively slow, but [mainland enterprises] aiming at business expansion overseas would first consider Hong Kong as their testing zone.”*

This kind of positioning is evident in many respondents' replies, mentioning the EU's General Data Protection Regulation and the mainland's Personal Information Protection Law. Although these laws are not applicable in Hong Kong, local enterprises inevitably look for places with the latest regulations when developing their products or even use them as testing grounds for gaining customer feedback on the market.

Some respondents considered the role of the government to be indispensable. Some believed that the government needs to take the lead in legislating and authorizing enforcement agencies to oversee institutions and enterprises. Only then will institutions and enterprises have the motivation to face up to problems and actively participate. Others held a more moderate view that the government should provide a platform for different organizations (including enterprises, universities, and the public sector) to participate and develop a framework and a flexible set of basic guidelines. On the whole, enterprises hoped that Hong Kong can have a common set of basic standards that can be fine-tuned to accommodate differences, rather than muddling through the old ways. Opinions suggested that the new consensus and standards should encourage data sharing across legal regions. In particular, the scale of local data is inherently smaller than that of other places, and the ambiguity of privacy standards discourages data sharing. This has led to a lack of data desperately needed by IT projects and AI start-ups. These issues cannot be solved solely by relying on more capital investment and talent.

### 2.2.3 Summary: Views on governance

Regarding the scope of concern around AI, the majority of respondents believed that experts are at the core of governance. They all believed in expert-led communication. This idea can be subdivided into two categories. One category believes that the work of education and interpretation is carried out unidirectionally under the definition of experts or authoritative figures, and laypersons who lack technical knowledge can only be passively educated. The other category emphasizes the role of the general public in interactively building a more complete consensus together with these experts. The former tends to believe that experts understand the needs of technology better than the general public, and effective communication lies in communicating the authenticity of these needs. The latter tends to think that demand is co-created and established by consensus. At the same time, many respondents believed that external third-party professionals, including universities and public institutions (such as statutory bodies that are not directly under the government), play an important role in explaining and educating the public on AI applications.

Hong Kong's positioning and open-market characteristics mean that local I&T enterprises are affected by global markets and other regional regulations. With the gradual development of different norms around the world, enterprises hoped that Hong Kong can also develop a common set of basic standards, in which the role of the government is considered essential.

## Ethical values and principles in artificial intelligence: Questionnaire results

In March 2022, the research team commissioned the Centre for Communication and Public Opinion Surveys (CCPOS) to conduct a telephone survey on ethical values and principles in AI governance from late April to early May, with a random sample size of 500 (see Appendix B).

### 3.1 High ranking of universally endorsed ethical values

We have identified five commonly mentioned ethical values in the literature and interviews and we asked respondents to rank them one by one based on perceived importance without hinting to them that the values may be in conflicts and may involve some trading-off. These values are "transparency", "personal freedom", "privacy", "robustness", and "unbiased". Regardless of gender, age, and education level, respondents generally agreed on the importance of these values in AI applications. About 40-50% of the respondents ranked each value as the most important (7 points) in the absence of a framework and specific guidance, and about 70-80% believed each value was important (5 points or higher). This suggests that respondents tended to rank these values high when the values were considered independently.

Respondents' responses to different ethical values do not differ significantly, and the preliminary ranking of values in terms of their importance (5 points or more) is as follows:

*Privacy > Robustness > Individual Freedom > Transparency > Unbiased*

### 3.2 The clustering of ethical values under trade-offs

Respondents were further asked to make trade-offs and retain only three of the above five ethical values as more important than the other two. Surprisingly, the value ranking obtained through value trade-offs only barely differs from the previous value ranking. Most respondents reserved individual freedom, privacy, and transparency as the top three values, while "unbiased" was the least important value that respondents chose to retain. Nevertheless, it is worth noting that clustering can be seen among these values. One group of respondents would jointly retain "individual freedom", "privacy", and "transparency", while some would jointly retain "robustness" and "unbiased". This reflects that respondents saw the values in the same paired choice as more relevant to each other. At the same time,

when the same respondent chose the highest score in "transparency", "privacy", or "individual freedom" (7 points), they also gave higher priority to "privacy" or "individual freedom" while giving lower priority to "transparency".

### **3.3 Value trade-offs contextualized under various technology types and scenarios**

In the third category of questions, we presented the following four "technology types and scenarios" to respondents (see Appendix C for details) so as to contextualize their responses:

#### **Scenario 1: Health Code System**

In order to combat the spread of the coronavirus, some countries introduce mobile technology, AI relationship analysis, and big data to assign "health codes" of different colours to display the health status of citizens, so as to control travel flow and break the chain of potential spread of the virus.

How health codes are generated by AI algorithms are not intelligible to citizens and the operation of the system may lack transparency.

#### **Scenario 2: Fraud Detection System**

To combat serious fraud, the government has developed a big data detection system that analyzes relevant data, such as tax status, health insurance, and other personal information, to generate fraud risk reports that help investigate fraud.

#### **Scenario 3: Driverless Autonomous Vehicles**

Let's say you are in a driverless autonomous car and there is only one person in the car. As the car drives into the tunnel, there is a pedestrian at the entrance to the tunnel. The self-driving car has only two options: either goes straight and hits the pedestrian or takes a turn and hits the tunnel wall, which would kill you. How would you choose between the options?

#### **Scenario 4: Robot Dog Surveillance**

The police departments of the New York City and Singapore have deployed robot dogs in public areas, which monitor the public and remind people to maintain social distancing during the epidemic through built-in microphones, and further notify the police if action is

required for violations. However, public reactions vary. New York City citizens tend to oppose it, while Singaporeans tend to support it.

In the first three scenarios, the respondents were asked to make trade-offs between two values (e.g., "unbiased" and "privacy"). We find that:

- System effectiveness (or "robustness") is less important when it comes to the trade-off against other values (for example: "transparency" and "privacy").
- Unlike the responses to the decontextualized question, there is a shift in the ranking of ethical values in AI – more respondents believed that system "transparency" is more important than "individual freedom". If we understand "effectiveness" as "robustness" (for example: the safety of "Driverless Autonomous Vehicles"), more respondents believed that "individual freedom" is more important than "robustness":

*Transparency > Individual Freedom > Effectiveness (or "Robustness")<sup>37</sup>*

- In the "Fraud Detection System" scenario, there is a stronger tendency for the public to think that system "transparency" and "individual freedom" are important (as compared to effectiveness) when compared with the other two scenarios ("Health Code System" and "Driverless Autonomous Vehicles")
- With regard to the "Health Code System", more respondents thought it was "difficult to choose" than in the "Fraud Detection System" scenario when they were asked to carry out value trade-offs:
  - The complexity of real-world experiences may make it harder for respondents to make a choice.
  - In the "Driverless Autonomous Vehicles" version of the trolley problem<sup>38</sup>, with the respondent as by default a passenger in an autonomous car in a traffic accident, faced with only two choices: hit and kill a pedestrian or turn and crash into the tunnel wall and kill himself, about half of the respondents chose the latter.

---

<sup>37</sup> Robustness is a technical concept describing a system's correct and error rate under a dynamic environment. Effectiveness measures only the correct and successful cases to an objective, without considering the economic value and efficiency of a system. Since the concept of robustness may be rather complex and difficult to be explained in scenarios 1 & 2 during our telephone survey, we opt for a simpler concept for them. The relation of the concepts could be understood as "what is not effective is not robust".

<sup>38</sup> Trolley Problem is a thought experiment originally proposed by philosopher Philippa Foot in "The Problem of Abortion and the Doctrine of Double Effect" in 1967. Variations of the Trolley Problem were subsequently designed to argue for and illustrate different ways of ethical decision-making.

### 3.4 Characteristics of acceptance and response to technology types and scenarios

After the respondents were provided with information and a framework for various application technologies and scenarios, they were asked to:

1. Determine whether to deploy specific AI applications (i.e., Health Code System, Fraud Detection System, and Driverless Autonomous Vehicles), and;
2. Indicate their level of acceptance towards a particular AI application (i.e., robot dog surveillance)

In terms of deployment, respondents expressed mixed opinions on the "Health Code System" and the "Fraud Detection System". Half of the respondents believed that they should be deployed, while the other half believed they should not. However, about two-thirds of the respondents said that driverless autonomous vehicles should not be deployed.

In terms of acceptance, 20.8% of the respondents found it "difficult to decide" on the robot dog surveillance application. Slightly more people tended not to accept it, accounting for 44.3%, while 34.2% were inclined to accept it.

The respondents' responses to the "Health Code System" and the "Fraud Detection System" are quite consistent, with about 80% of the respondents expressing the same opinion on both types of AI applications: When they thought one of the applications should be adopted, there is a good chance that they thought the other should also be adopted, and vice versa. If "difficult to decide" is excluded and only accepted or not accepted are considered, respondents' views on "robot dog surveillance" are also very similar to these two types of applications.

However, in the case of "Driverless Autonomous Vehicles", respondents expressed opinions that are not consistent with those of the other three types of intelligent systems. We cannot predict respondents' opinions on the other three types of intelligent systems based on their opinions on whether or not to adopt autonomous vehicles.

Regression analysis with demographic data reveals that for the "Health Code System" and the "Fraud Detection System", older respondents were more likely to accept them. Conversely, the higher the level of educational attainment, the less inclined they were to

accept them. With regard to "Robot Dog Surveillance", only older respondents were more willing to accept the application.

Contrary to the other findings, in the "Driverless Autonomous Vehicles" scenario, respondents in our study were more likely not to accept it as they aged, while the higher the level of educational attainment they had, the more inclined they were to accept the application.

### **3.5 Governance: Whose views should be prioritized?**

The term governance has many meanings. Yet, one thing in common among different usages of the term is that governance is not uniquely governmental. Structures, corporations, and individuals outside the government often play important roles in governance. Therefore, when we ask "whose views should be prioritized", the question partly reflects how different roles in governance should be allocated.

After the respondents were presented with the value trade-off questions for specific application technologies and scenarios, they were asked who they believed was the best person to address these situations. Nearly 60% of respondents believed that the opinions of "affected individuals" should be given priority regardless of technology application scenarios. Only about 2% of the respondents believed that private businesses are the best party to address these trade-off situations.

Respondents' responses regarding the "Health Code System" and the "Fraud Detection System" are quite consistent. Apart from the affected public, "relevant public bodies" are the most widely chosen (about one-quarter to one-fifth of the respondents), followed by "third-party professionals" (about one-eighth of the public). In the scenario of "Driverless Autonomous Vehicles", the results are slightly different, with more people choosing "third-party professionals" (about one-quarter) and less than 10% choosing "relevant public bodies".

In addition, 50-65% of the respondents who chose "difficult to choose" under the value trade-off scenario still tended to believe that "affected individuals" are the best persons to address these trade-off situations.

All in all, the following is the order of priority given by respondents:

*Affected citizens > Relevant Public Bodies > Third-Party Professionals >> Private Businesses (the least selected category with less than 2% of the respondents choosing)*



## 4. Discussion and policy advice

In this chapter, we will select some of the most notable interviews and questionnaire results and compare them with other domestic and foreign research results for interpretation.

### 4.1 Contextualization matters

According to foreign literature, Americans have mixed feelings about AI technology. Our interview respondents shared a similar attitude. They neither showed entirely negative nor entirely positive reactions. It is also noteworthy that the ethical values or principles that were seen to deserve particular attention in AI technology mentioned in the interviews do not differ much from the rest of the world. For example, both interview and questionnaire results show that people in Hong Kong value privacy and individual freedom, which is consistent with the media coverage of the smart lamppost incident.

It is worth noting that the questionnaire results show that the general public tends to rank highly the values or principles in AI ethics when they are mentioned in a decontextualized way. People tend to ignore possible conflicts among these values or principles. Such tendency may serve as a convenient public consensus to fuel policy discussions. Research of public perception of AI in the US and the European Union region mentioned in section 1.5 serves as a signal influencing policy decisions for democratic governments. In an active civil society, citizens will likely adopt regulatory mechanisms that derive from their claim right, quoting the relevant values/principles and the appeal to the judiciary to secure these rights. The tendency mentioned above encourages citizens and civil societies to express and strive for their rights proactively, creating a force that pushes the judiciary to execute relevant legal mechanisms. For emerging topics like AI ethics and governance, such tendency can easily attract public attention and sympathy, which may be helpful to drive public discussions and create policy demand.

However, we need to consider whether the same strategy is applicable in Hong Kong. During our interviews, many interviewees expressed confusion about the problem of decontextualized value ranking and showed skepticism and reservations about the tendency. In the absence of a specific context, the tendency to rank the values or principles in AI ethics highly provide limited practical meaning for policy making. It is more likely to mislead the public into thinking that these values can be met unconditionally and without trade-offs. Especially when there is little difference in the ranking of importance between individual values in different contexts, the tendency of uncritically assigning high priority to ethical values and principles can easily lead us into an intractable contradiction and

antagonistic impasse. The survey results show that the majority of the public expressed that it was "difficult to choose" when confronted with value trade-offs and believed that "affected individuals are the best candidates to make priority decisions to solve the situation". This shows that even if the tendency is a consensus, it is not an actionable consensus for Hong Kong, at least at this stage.

***Policy Advice (1):***

To ensure that ethical values or principles are understandable, it is essential to frame them in specific contexts or scenarios of AI applications, instead of merely presenting abstract terms of ethical values or principles.

## **4.2 Framework on the application-scenario type and cultural differences**

We believe that instead of uncritically assigning high priority to values or principles in AI ethics in a decontextualized way and establishing a one-size-fits-all normative standard for AI applications, it is better to understand how these principles or values are generated and then establish standards accordingly. Our questionnaire results reveal that citizens' perception of value differs greatly across various scenarios of AI applications, and the underlying reasons require careful examination. For instance, nearly two-thirds of the respondents did not accept driverless autonomous vehicles. When asked about the trade-off between increasing restrictions on ordinary people to improve system robustness and freedom of movement of ordinary people, more respondents tended to choose the latter. At the same time, our interviewees emphasized that safety and robustness are crucial for the driverless autonomous vehicle systems.

To address the concerns of the public, specific measures should thus be proposed to address the need for safety and robustness when this technology is introduced. In other words, Hong Kong may need more testing grounds for driverless autonomous vehicles to help improve safety and robustness. In addition, the technology should be applied to closed systems first, such as airport ground transport, railways, etc., so as to reduce the impact on other road users. For the "Health Code System" and the "Fraud Detection System", both questionnaire respondents and interviewees agreed that transparency is essential. For such applications, applying the operation in a closed system may not serve the purpose. Fundamentally, the two types of smart applications satisfy different needs and have different content focuses, such as security and transparency. Over-generalizing different intelligent applications is thus not appropriate.

We would particularly wish to highlight that, in addition to integrating a common framework to distinguish between different types and scenarios of AI technologies, it is also necessary to be aware of differences in social or demographic contexts. For example, some studies have analyzed the correlation between the demographic data of the Americans and the acceptance of artificial intelligence and found that the acceptance has a certain correlation with gender, education, age, race, political orientation, etc. However, in our study, we have only observed an association between acceptance, and education and age in Hong Kong, but no association was found with gender differences.

Therefore, in the case of "Driverless Autonomous Vehicles", communication management and education should target older and less educated people, while for the "Health Code System" and the "Fraud Detection System", it should be the other way round. In other words, different interest groups are present in different societies with different perceptions towards similar intelligent applications. Early research has pointed out that differences in individual ethical concerns of different countries may still exist but in more specific forms. In this study, we offered a limited number of intelligent applications and scenarios on a non-experimental basis. Experiments on a larger scale with cross-cultural frameworks would be necessary to explore where to find common grounds while acknowledging the differences.

#### ***Policy advice (2):***

Instead of offering a one-size-fits-all framework, special attention should be paid to cater for cultural and political diversity in the establishment of application-specific or sector-specific norms, while applying them incrementally to the actual situation in Hong Kong. Policymakers can make reference to the idea of "relevant markets" in the latest Competition Ordinance in Hong Kong or the sector-specific approach in the government competition policy in the 1990s-2000s.

### **4.3 Explaining distrust towards enterprises**

While our interviewees did not show a general distrust towards enterprises, data from our survey respondents tells a different story. There was a general distrust towards enterprises from the public (as discussed in section 3.5) and it is similar to the situation in the US and EU countries. Combining the results of the questionnaire with the opinions of the interviewees, we have identified two possible reasons for this:

(1) Enterprises do not have incentive to promote good governance: They prioritize their own business interests over those of citizens or consumers, and even if their acts pose a negative impact on the society, they seldom face any penalty.

(2) Enterprises do not have sufficient capacity to promote good governance: The capabilities of an enterprise are derived from its consumers, and the trust that individual enterprises gain comes only from their own consumers, instead of from the whole public. Competition between businesses (such as between small-/medium-sized start-ups and large corporations) and the lack of clear and consistent needs among consumer groups make it difficult for enterprises to reach uniform standards or consensus.

To address the first reason, it is important to have government regulations and laws in place. For the second reason, it is important to increase external involvement and assistance from public institutions and third-party professionals, including statutory bodies, universities, and audit institutions. However, enterprises still have an important role to play. While our questionnaire results show that the views of the affected public should be foremost considered, this does not mean that the views of enterprises, governments, and professionals are not important. The Smart City Blueprint is a government-led initiative that introduces many AI technologies into governmental services. We should not forget that many of these projects involve corporate collaboration. Free market competition makes enterprise one of the major driving forces for AI innovation -- especially in the US. Ignoring views from enterprises from the governance formula would diminish their incentives and flexibility in technology innovation, leading to undesirable technology stagnation which should be avoided.

### ***Policy advice (3)***

There should be authorized agencies to research, review, coordinate, and enforce standards and norms on AI applications. The Office of the Government Chief Information Officer (OGCIO) has provided a template for Application Impact Assessment, a kind of Impact Assessment Report (IAR), in September 2022<sup>39</sup>. It would be advisable to make it mandatory and to require organizations to introduce independent external expert advice, auditing, or supervision on their AI projects in future for better transparency and governance. The government can also consider requiring all AI applications implemented by listed companies in the Hong Kong Stock Exchanges to be assessed for governance in their ESG reporting for public scrutiny.

---

<sup>39</sup> Office of the Government Chief Information Officer. (2022). *Ethical artificial intelligence framework*. Office of the Government Chief Information Officer. [https://www.ogcio.gov.hk/en/our\\_work/infrastructure/methodology/ethical\\_ai\\_framework/](https://www.ogcio.gov.hk/en/our_work/infrastructure/methodology/ethical_ai_framework/)

For high-stakes application types and scenarios (e.g., autonomous vehicles and financial services), an incremental and gradual approach should be adopted with constant review instead of following a rigid framework. This will enable space for enterprises and individuals to explore, develop, and adapt to market-oriented standards or performance pledges for AI applications on a voluntary basis, which can help amass professional advice from the society.

#### 4.4 Meaning and strategies of educational promotion and communication

In section 2.2 we have discussed the different views of respondents with technical and management backgrounds on the role of experts. Those with technical backgrounds tended to believe that non-experts lacked the technical knowledge for understanding ethical issues related to technology. While there is no doubt about the technical knowledge of experts, it is problematic to equate technical knowledge with ethics of technology. For instance, ethical standards for embryo research are often linked to scientific and technological knowledge in mutual monitoring between researchers. Technical knowledge is useful to ethical behaviors only after clear norms, standards, and governance mechanisms are ready. However, our current norms and governance mechanisms for AI applications are not ready.

When comparing the opinions obtained from the interviews with the survey results in section 3.5, there is no data to support the idea that Hong Kong citizens specifically agree with the notion that "third-party professionals are the best candidates to make decisions on the priority situation". It may be more practical and preferable not to equate technical knowledge with technology ethics or technology literacy. Rather, it would be better to regard the latter as a consensus jointly established by experts and the public.

Regarding increasing public recognition of AI, we believe that distinguishing between different types of intelligent applications and scenarios is more necessary than simply promoting technical knowledge. Other studies have shown that Hong Kong citizens are likely to have less trust in AI technology as their understanding of it increases<sup>40</sup>.

Respondents in a study by Salesforce were asked to self-assess their understanding and trust in intelligent applications. The conclusions of the study partially echo the findings of our interviews, namely the asymmetry between functionality and technical details among citizens in section 2.1, that is, when citizens do not understand the functions or needs of technology, they would focus more on the application of AI technology, leading to speculation, suspicion, and mistrust. However, we have some reservations about the

---

<sup>40</sup> Salesforce. (2018). *Artificial intelligence in Asia: trust, understanding and the opportunity to re-skill*.

study's conclusions for Hong Kong because respondents' self-assessment of their understanding of smart technology does not necessarily correspond to their actual understanding of AI technology.

Salesforce's study also finds that in most other Asian regions, including India, Singapore, Malaysia, Thailand, Indonesia, and the Philippines, respondents' self-assessment of their understanding of AI technologies positively correlated with trust in technology. This finding is more in line with the views of our interviewees that increased understanding of AI technology will reduce distrust of technology. Our questionnaire provided specific scenarios for several different intelligent applications. For respondents with higher education attainment, they have a higher degree of trust in autonomous vehicles, and vice versa. After completing the relevant situational questions, we expected the respondents to have a more specific understanding of intelligent applications. After that, when we asked respondents to self-rate whether their acceptance of AI changed after completing the question, there was a slight overall increase in their acceptance level<sup>41</sup>.

More importantly, when we conducted demographic regression analysis on different types of AI applications (section 2.4), it was found that respondents tended to regard the "Health Code System" and the "Fraud Detection System" as belonging to the same category and "Driverless Autonomous Vehicles" as another. Moreover, expecting consistent observations between demographic factors and acceptance levels of AI without distinguishing different types of AI can be misleading. For instance, in our study higher education level and younger age negatively correlated with the acceptance of the "Health Code System" and the "Fraud Detection System" and yet positively correlated with the acceptance of "Driverless Autonomous Vehicles". Therefore, **it is necessary to distinguish between different types and scenarios of intelligent applications, identify individual groups of interest, and use different approaches to address different reasons for skepticism and distrust, rather than indiscriminately promoting technical knowledge and uncritically assigning high ranking to values or principles in AI ethics.**

#### *Policy advice (4)*

Technology literacy should be improved by addressing different types and scenarios of AI applications through consultations and other consensus-building strategies. Trust and trustworthiness are different. It is possible for a trustworthy AI project not to be trusted. Expert-led discussions and works on AI ethics without good public engagement may still

---

<sup>41</sup> Refers to Q.21 of the questionnaire in Appendix C: 68.3% of the respondents reported no change to their acceptance level to AI application, 21% reported higher level of acceptance, 9.1% reported lower, and 1.5% refused to answer.

be insufficient as expert knowledge may be highly reliable with regard to a target outcome, but still, the outcome will not get enough support from the public. Support research from third-party professionals, including statutory bodies and universities (which are independent from the government), would be needed to provide independent, evidence-based policy advice, offering perspectives different from the rigid, execution-oriented routines.

## 4.5 Conclusion

Uncritically assigning high priority to values or principles in AI ethics can easily lead us into intractable dilemmas and confrontational impasses, which do not guide us to actionable consensus. We believe that, in addition to proposing frameworks to distinguish between different types and scenarios of AI technologies, it is also necessary to be mindful of sociocultural or demographic diversity. Experimental research taking cross-cultural backgrounds into account is needed to guide us towards finding a common ground for inclusiveness and diversity.

Different parties should be mobilized for norm-building as there is no single party that has a wide endorsement from the public and also possesses technological expertise. Legal intervention and government regulation are advisable for norm-building in Hong Kong, especially mandatory reporting. At the same time, it is important to increase the external involvement and assistance of public institutions and third-party professionals, including statutory bodies, universities, and audit institutions. These parties can work alongside with technology enterprises in AI governance.

To increase citizens' recognition of AI, we believe that distinguishing between different types of intelligent applications and scenarios is more fundamental than simply promoting technology literacy. In particular, it may be more practical and desirable to regard technology ethics as a consensus jointly established by experts and citizens.

## References

- Algorithm Watch. (2020). *AI ethics guidelines global inventory*. <https://inventory.algorithmwatch.org/>
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, 563(7729), 59-64. <https://doi.org/10.1038/s41586-018-0637-6>
- Bevir, M. (2009). *Key concepts in governance*. SAGE.
- Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). *Notes from the AI frontier: applications and value of deep learning*. <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>
- European Commission. (2017). *Special Eurobarometer 460: attitudes towards the impact of digitisation and automation on daily life*. European Commission.
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI*.
- 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>
- Institute for Management Development and Singapore University for Technology and Design. (2021). *Smart city index 2021*. Institute for Management Development. [https://imd.widen.net/s/x25tgd8drv/smart\\_city\\_index\\_2021](https://imd.widen.net/s/x25tgd8drv/smart_city_index_2021)
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399. <https://doi.org/10.1038/s42256-019-0088-2>
- Jobin, A., Man, K., Damasio, A., Kaissis, G., Braren, R., Stoyanovich, J., Van Bavel, J. J., West, T. V., Mittelstadt, B., Eshraghian, J., Costa-jussà, M. R., Tzachor, A., Jamjoom, A. A. B., Taddeo, M., Sinibaldi, E., Hu, Y., & Luengo-Oroz, M. (2021). AI reflections in 2020. *Nature Machine Intelligence*, 3(1), 2-8. <https://doi.org/10.1038/s42256-020-00281-z>
- Kostka, G., Steinacker, L., & Meckel, M. (2021). Between security and convenience: facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. *Public Understanding of Science*, 30(6), 671-690. <https://doi.org/10.1177/09636625211001555>
- Müller, V. C. (2021). *Ethics of artificial intelligence and robotics*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/>
- MacCarthy, M., & Propp, K. (2021, May 4). *Machines learn that Brussels writes the rules: The EU's new AI regulation*. Brookings. <https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/>
- Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature (London)*, 568(7751), 163-165. <https://doi.org/10.1038/d41586-019-01099-5>
- Moar, J., & Bainbridge, M. (2022). *Smart cities: key technologies, environmental impact and market forecasts 2022-2026*. J. Research. <https://www.juniperresearch.com/researchstore/healthcare-government/smart-cities-research-report>
- Northeastern University and Gallup. (2018). *Optimism and anxiety: views on the impact of artificial intelligence and higher education's response*. Gallup. <https://www.northeastern.edu/gallup/pdf/OptimismAnxietyNortheasternGallup.pdf>



- OECD. (2022). *Recommendation of the Council on artificial intelligence* (OECD/LEGAL/0449).  
<https://legalinstruments.oecd.org/api/print?ids=648&lang=en>
- Office of the Government Chief Information Officer. (2022). *Ethical artificial intelligence framework*. Office of the Government Chief Information Officer.  
[https://www.ogcio.gov.hk/en/our\\_work/infrastructure/methodology/ethical\\_ai\\_framework/](https://www.ogcio.gov.hk/en/our_work/infrastructure/methodology/ethical_ai_framework/)
- Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., & Niebles, J. C. (2019). *The AI index 2019 annual report*. Stanford University.  
[https://hai.stanford.edu/sites/default/files/ai\\_index\\_2019\\_report.pdf](https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf)
- Pew Research Center. (2022). *AI and human enhancement: Americans' openness is tempered by a range of concerns*. Pew Research Center. [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS\\_2022.03.17\\_AI-HE\\_REPORT.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS_2022.03.17_AI-HE_REPORT.pdf)
- Pierre, J., & Peters, B. G. (2005). Governance: A Garbage Can Perspective. In *Governing complex societies: trajectories and scenarios* (pp. 49-63). Palgrave Macmillan UK. [https://doi.org/10.1057/9780230512641\\_3](https://doi.org/10.1057/9780230512641_3)
- Salesforce. (2018). *Artificial intelligence in Asia: trust, understanding and the opportunity to re-skill*.
- Sartori, L., & Bocca, G. (2022). Minding the gap(s): public perceptions of AI and socio-technical imaginaries. *AI & society*.  
<https://doi.org/10.1007/s00146-022-01422-1>
- Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., Lyons, T., Etchemendy, J., Grosz, B., & Bauer, Z. (2018). *The AI index 2018 annual report*. Stanford University.  
[https://hai.stanford.edu/sites/default/files/2020-10/AI\\_Index\\_2018\\_Annual\\_Report.pdf](https://hai.stanford.edu/sites/default/files/2020-10/AI_Index_2018_Annual_Report.pdf)
- Surya, L. (2017). AI in information technology and its future in the United States. *International Journal of Creative Research Thoughts (IJCRT)*, ISSN, 2320-2882.
- The Guardian. (2019). Hong Kong riot police beat protesters at anti-surveillance rally. *The Guardian*.  
<https://www.theguardian.com/world/2019/aug/24/hong-kong-fresh-rallies-as-protesters-target-airport-transport>
- Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, 2(1), 10-12. <https://doi.org/10.1038/s42256-019-0136-y>
- Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking artificial intelligence principles. *arXiv preprint arXiv:1812.04814*.
- Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. University of Oxford.  
[https://isps.yale.edu/sites/default/files/files/Zhang\\_us\\_public\\_opinion\\_report\\_jan\\_2019.pdf](https://isps.yale.edu/sites/default/files/files/Zhang_us_public_opinion_report_jan_2019.pdf)

## Appendices

## Appendix A: Interview Protocol Outline

The AI Governance working group developed a semi-structured interview protocol to be used in formal interviews as part of its research proposal. This protocol provides (1) introductory remarks; (2) two general, non-directive open-ended questions on Hong Kong's smart city initiatives and public sector data governance; and (3) value-based structured questionnaire and application-specific scenarios.

### Introductory Protocol

Thank you for agreeing to participate in this interview. As we described in our invitation letter, this interview is part of a collaborative research project led by scholars from the City University of Hong Kong and the University of Hong Kong. The purpose of the research is to identify areas of concern and a governance framework related to the application of AI technology in Hong Kong.

Before we begin this interview, I would like to confirm that you are participating on a voluntary basis.

[If no: thank the interviewee, then stop the interview immediately.]

[If yes: continue]

If you prefer, you will not be cited by name. Your responses will be anonymized so that any finding of the project will not be traced back to you. What is your preference?

[Mark down the interviewee's preference: Anonymize / Not Anonymize ]

Finally, we would like to record today's interview. We will use the recording to confirm the interview notes, such as by helping us to capture exact phrases. Is this okay?

[If no: Okay, we will not record today's interview.]

[If yes: Thank you; we will begin recording now.]

In case of non-anonymized, would you prefer us to send the quotation of your comments/ responses for verification before finalization the report.

[If no: it is not needed.]

[If yes: it is needed.]

.....

### General Open-ended Questions

#### ***Smart City and Automation (10 minutes)***

Let us begin with a broad application of AI today in Smart City initiative and automation.

- When you think of the applications of artificial intelligence in the Smart City initiatives, what comes to your mind?

- In your opinion, what composes our Smart City?
- In comparison to the traditional means, how the AI technology may make a difference?
  - [If not mentioned] It is believed that in the Smart City initiative, public data will be collected, verified, analyzed, and used to offer services by automated machines. In comparison to doing these tasks manually, what are the differences?
  - [If not mentioned] What benefits come to your mind?
  - [If not mentioned] What risks and challenges come to your mind?

**Public Sector Data\* (10 minutes)**

Technology changes how we treat our data. Now we would like to focus on how the public sector handles data in developing AI applications.

- In succession to our previous discussion, suppose public data will be collected, stored, and used by automated means, does the public need a different way to manage the big data environment?
- [If not mentioned] In our Smart City, what is your opinion on the idea that certain types of data be required for mandatory collection and/or disclosure (e.g. taxation-related information by authority)?
- [If not mentioned] What is your opinion on the right of the certain data of data subject (e.g. right to be forgotten) after a certain time period?

.....

**Questionnaire and Vignettes on Application-specific Scenarios**

The next part of this interview will be on a more structured survey on ethical values and principles over governance of AI technology, followed by 3 draft vignettes. We created a set of 3 vignettes that are meant to help decision-makers visualize a range of implications of AI.

[Show questionnaire “*Values/Principles in AI Governance: Scenarios of Application*”]

**(10-20 minutes)**

**Vignette Closeouts**

Now that you have considered a few vignettes, are there any others you would like us to consider? We would be happy to either talk about these now or receive your thoughts over email.

.....

**Close Out and Next Steps (5 minutes)**

Thank you for speaking with us today. Before we end, there is one more question we would like to ask --

- Are there any candidate interviewees whom you recommend we should speak with? We would be happy to either talk about these now or receive your thoughts over email.

## Appendix B: 電話訪問：樣本資料

### 「人工智能治理中的倫理價值與原則」意見調查

#### 調查概況

調查方法	以隨機抽樣的方式，利用電腦輔助電話訪問系統，由訪問員進行電話訪問																																		
調查日期	2022年4月25-29、5月3-6、10-11日(共11天)																																		
調查對象	18至65歲，能操粵語的香港市民																																		
抽樣方法	<p><b>固網電話：</b> 將通訊事務管理局已經發出的固網電話號段，做成一個電話號碼庫。從這個電話號碼庫中隨機抽出號碼，並跟據過往經驗盡量剔除無效號碼。 成功接觸住戶後，若其家中有多於一位合資格接受訪問的成員，便會採用「即將生日」的方式來選出一位最快到達生日日期的合資格成員進行訪問。</p> <p><b>手提電話：</b> 將通訊事務管理局已經發出的手提電話號段，做成一個電話號碼庫。從這個電話號碼庫中隨機抽出號碼，並跟據過往經驗盡量剔除無效號碼。</p>																																		
電話號碼的處理方法	<p>撥出的電話號碼若遇無人接聽、線路繁忙，或合適的被訪者不在家等，電腦系統會安排訪問員在不同日期或時段再次致電。</p> <p><b>固網電話：</b>致電三次皆未能成功訪問，便會停止致電該號碼；但若有兩次被拒絕，便會立刻停止致電。</p> <p><b>手提電話：</b>致電三次皆未能成功訪問，便會停止致電該號碼；但若有一次被拒絕，便會立刻停止致電。</p>																																		
樣本數目	510 個成功個案																																		
抽樣誤差	± 4.3% 以內 (可信度設於 95%；即代表有 95% 信心，百分比誤差會在這範圍以內。)																																		
回應率	36%																																		
<p>以下為撥出電話號碼的情況，以及回應率的計算方法：</p> <table border="1"> <tr> <td>撥出的電話號碼總數</td> <td>41128</td> </tr> <tr> <td><b>A. 確定不適合訪問的電話號碼數目 (Ineligibles)</b></td> <td>24798</td> </tr> <tr> <td>A1. 無效號碼</td> <td>23403</td> </tr> <tr> <td>A2. 非住宅 / 商業電話</td> <td>284</td> </tr> <tr> <td>A3. 傳真 / 數據機 / 傳呼機</td> <td>595</td> </tr> <tr> <td>A4. 無合適被訪者</td> <td>516</td> </tr> <tr> <td><b>B. 未確定有沒有合適被訪者的電話號碼數目 (Unknown)</b></td> <td>15451</td> </tr> <tr> <td>B1. 無人接聽</td> <td>7393</td> </tr> <tr> <td>B2. 線路繁忙</td> <td>2178</td> </tr> <tr> <td>B3. 密碼阻隔</td> <td>22</td> </tr> <tr> <td>B4. 語言不通</td> <td>25</td> </tr> <tr> <td>B5. 掛線前仍未確定為住宅或合適被訪者</td> <td>5736</td> </tr> <tr> <td>B6. 長途電話</td> <td>97</td> </tr> <tr> <td><b>C. 確定有合適被訪者的電話號碼數目 (Eligibles)</b></td> <td>879</td> </tr> <tr> <td>C1. 拒絕訪問 (包括訪問中途拒絕)</td> <td>305</td> </tr> <tr> <td>C2. 合適的被訪者未能在調查期間接受或完成訪問</td> <td>64</td> </tr> <tr> <td>C3. 成功訪問</td> <td>510</td> </tr> </table>		撥出的電話號碼總數	41128	<b>A. 確定不適合訪問的電話號碼數目 (Ineligibles)</b>	24798	A1. 無效號碼	23403	A2. 非住宅 / 商業電話	284	A3. 傳真 / 數據機 / 傳呼機	595	A4. 無合適被訪者	516	<b>B. 未確定有沒有合適被訪者的電話號碼數目 (Unknown)</b>	15451	B1. 無人接聽	7393	B2. 線路繁忙	2178	B3. 密碼阻隔	22	B4. 語言不通	25	B5. 掛線前仍未確定為住宅或合適被訪者	5736	B6. 長途電話	97	<b>C. 確定有合適被訪者的電話號碼數目 (Eligibles)</b>	879	C1. 拒絕訪問 (包括訪問中途拒絕)	305	C2. 合適的被訪者未能在調查期間接受或完成訪問	64	C3. 成功訪問	510
撥出的電話號碼總數	41128																																		
<b>A. 確定不適合訪問的電話號碼數目 (Ineligibles)</b>	24798																																		
A1. 無效號碼	23403																																		
A2. 非住宅 / 商業電話	284																																		
A3. 傳真 / 數據機 / 傳呼機	595																																		
A4. 無合適被訪者	516																																		
<b>B. 未確定有沒有合適被訪者的電話號碼數目 (Unknown)</b>	15451																																		
B1. 無人接聽	7393																																		
B2. 線路繁忙	2178																																		
B3. 密碼阻隔	22																																		
B4. 語言不通	25																																		
B5. 掛線前仍未確定為住宅或合適被訪者	5736																																		
B6. 長途電話	97																																		
<b>C. 確定有合適被訪者的電話號碼數目 (Eligibles)</b>	879																																		
C1. 拒絕訪問 (包括訪問中途拒絕)	305																																		
C2. 合適的被訪者未能在調查期間接受或完成訪問	64																																		
C3. 成功訪問	510																																		
<p>回應率的計算方法如下：</p> $\text{Completed} / [\text{Eligibles} + \text{Unknown} \times \text{Eligibles} / (\text{Eligibles} + \text{Ineligibles})]$ $= 510 / [879 + 15451 \times 879 / (879 + 24798)]$ $= 0.3622 \text{ (即 36\%)}$																																			

## Appendix C: Questionnaire 調查問卷

香港中文大學 傳播與民意調查中心

「人工智能治理中的倫理價值與原則」意見調查

2022 年 3 月

### 第一部份 抽樣及確定

#### 介紹

你好。呢度係 中文大學 傳播與民意調查中心 打嚟架，我哋做緊一項有關「社會應用人工智能」嘅意見調查，啲題目好簡單，麻煩你幫幫忙呀。

#### 抽樣

首先，我哋要喺你屋企隨機抽一位成員做訪問。  
請問你屋企，總共有幾多位 18 至 65 歲嘅成員呢？

【如果只有 1 位，便訪問這 1 位】

【如果多過 1 位，讀出「為咗隨機抽樣，我哋想訪問嚟緊最快生日嗰一位。」】

#### 確定

想確定一下，請問(先生/小姐/女士) 你依家係唔係 18 至 65 歲嘅香港居民呢？

1. 係
2. 唔係【讀出「唔好意思，我哋想訪問 18 至 65 歲嘅香港居民。」，重新抽樣】

性別(不用問)

1. 男
2. 女

## 第二部份 問卷題目

依家唔少地方都會應用 AI 人工智能幫助管治，但有機會涉及價值或者原則嘅問題。以下我哋會講一啲情景，想了解你嘅取向。

### 情境 1：健康碼系統

現時有部分國家為對抗冠狀病毒嘅傳播，利用移動科技技術、人工智能關係分析同大數據，分配不同顏色嘅「健康碼」，以顯示市民嘅健康狀況，從而管制市民嘅出入，用作打破潛在感染鏈。

呢啲由人工智能演算法分配嘅健康碼，市民無法理解當中運作，可能缺乏透明度。

Q1. 喺阻止病毒傳播嘅成效，同埋系統嘅透明度之間，你認為邊樣重要啲呢？你可以俾 1 至 7 分，愈接近 1 分代表系統透明度比較重要，愈接近 7 分代表系統成效比較重要，中間嘅 4 分代表難以取捨？1 至 7 分，你俾幾多分？

Transparency is more important 透明較重要 1			Hard to determine 難以取捨 4			Beneficence, do no harm and robustness are more important 行善、不傷害及穩健較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

Q2. 另外，執行健康碼會限制咗部份人嘅行動自由，從而降低病毒喺社區傳播嘅風險。喺阻止病毒傳播嘅成效，同埋個人行動自由之間，你認為邊樣重要啲呢？你可以俾 1 至 7 分，愈接近 1 分代表個人自由比較重要，愈接近 7 分代表系統成效比較重要，中間嘅 4 分代表難以取捨？1 至 7 分，你俾幾多分？

Freedom and autonomy are more important 自由及自主較重要 1			Hard to determine 難以取捨 4			Beneficence, do no harm and robustness are more important 行善、不傷害及穩健較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

**Q3.** 收集到高質量嘅數據可以幫助人工智能系統運作得更加準確，甚至可能達至無偏差，但市民個人隱私就缺乏保障，喺系統運作無偏差，同埋個人私隱之間，你認為邊樣重要啲呢？你可以俾 1 至 7 分，愈接近 1 分代表系統無偏差比較重要，愈接近 7 分代表個人私隱比較重要，中間嘅 4 分代表難以取捨？1 至 7 分，你俾幾多分？

Fairness and bias-free are more important 公平無偏差較重要 1			Hard to determine 難以取捨 4	9		Privacy is more important 私隱較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

**Q4a.** 你認為香港社會應唔應採用呢個「健康碼」系統？

1. 應該
2. 唔應該(跳至 Q5)
3. 無答案/拒絕回答(跳至 Q5)

**Q4b.** 咁面對「健康碼」呢個人工智能系統嘅道德困境，你認為應該優先考慮邊類人士既意見？係相關公共機構，包括政府機關及法定機構、私人企業、第三方專業人士、受影響嘅市民，定係其他？

1. 相關公共機構(包括政府機關及法定機構)
2. 私人企業
3. 第 3 方專業人士
4. 受影響嘅市民
5. 其他(請註明)
6. 無答案/拒絕回答



情景 2：欺詐偵測系統

為打擊嚴重欺詐罪案，政府開發大數據檢測系統，透過分析相關數據，例如稅務狀況、醫療保險及其他個人信息等，得出欺詐風險報告，有助於調查欺詐行為。

Q5. 政府唔會披露欺詐檢測系統嘅運作細節，亦唔會通知市民佢地嘅數據已被使用，以避免受外部操控而破壞系統嘅成效，但系統運作就會缺乏透明度。喺系統成效，同埋系統透明度之間，你認為邊樣重要啲呢？你可以俾 1 至 7 分，愈接近 1 分代表系統透明度比較重要，愈接近 7 分代表系統成效比較重要，中間嘅 4 分代表難以取捨？1 至 7 分，你俾幾多分？

Transparency is more important 透明較重要 1			Hard to determine 難以取捨 4			Beneficence, do no harm and robustness are more important 行善、不傷害及穩健較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

Q6. 政府利用欺詐檢測系統，可能有助減少社會欺詐罪行，有利於社會整體穩定。但同時檢測系統無可避免地會出現誤報案件，導致對少數人進行不必要嘅調查並限制佢哋嘅自由。喺系統成效，同埋個人自由之間，你認為邊樣價值重要啲呢？你可以俾 1 至 7 分，愈接近 1 分代表個人自由比較重要，愈接近 7 分代表系統成效比較重要，中間嘅 4 分代表難以取捨？1 至 7 分，你俾幾多分？

Freedom and autonomy are more important 自由及自主較重要 1			Hard to determine 難以取捨 4			Beneficence, do no harm and robustness are more important 行善、不傷害及穩健較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

Q7. 假設檢測系統成功識別出一啲欺詐者，並對潛在犯罪有威懾嘅效用。然而，由於無法解釋系統嘅運作方式，有可能影響司法嘅公正性。喺系統成效，同埋司法公正之間，你認為邊樣價值重要啲呢？你可以俾 1 至 7 分，愈接近 1 分代表司法公正比較重要，愈接近 7 分代表系統成效比較重要，中間嘅 4 分代表難以取捨？1 至 7 分，你俾幾多分？

Justice is more important 司法正義較重要 1			Hard to determine 難以取捨 4			Beneficence, do no harm and robustness are more important 行善、不傷害及穩健較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分                      8. 無答案/拒絕回答

Q8. 收集到高質量嘅數據可以幫助分析系統運作得更加準確，甚至可能達至無偏差，但市民個人隱私就缺乏保障，你認為邊樣重要啲呢？你可以俾 1 至 7 分，愈接近 1 分代表系統無偏差比較重要，愈接近 7 分代表個人私隱比較重要，中間嘅 4 分代表難以取捨？1 至 7 分，你俾幾多分？

Fairness and bias-free are more important 公平無偏差較重要 1			Hard to determine 難以取捨	9		Privacy is more important 私隱較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分                      8. 無答案/拒絕回答

Q9a. 你認為香港社會應唔應該採用呢個欺詐檢測系統？

1. 應該
2. 唔應該(跳至 Q10)
3. 無答案/拒絕回答(跳至 Q10)

**Q9b.** 咁面對「欺詐檢測系統」嘅道德困境，你認為應該優先考慮邊類人士既意見？係相關公共機構，包括政府機關及法定機構、私人企業、第三方專業人士、受影響嘅市民，定係其他？

1. 相關公共機構(包括政府機關及法定機構)
2. 私人企業
3. 第3方專業人士
4. 受影響嘅市民
5. 其他(請註明)
6. 無答案/拒絕回答

**情景3：無人駕駛自動駕駛汽車**

**Q10.** 你認為香港社會應唔應該採用無人駕駛自動汽車系統呢？

1. 應該
2. 唔應該
3. 無答案/拒絕回答

**Q11.** 假設你係一架無人駕駛自動汽車裏面，車上只有你1個人。喺你駕駛進入隧道之際，有一名行人喺隧道入口。自動駕駛汽車只有兩種選擇：一係繼續直行並撞死行人，二係轉向撞上隧道牆壁並會將你撞死。兩者之間，你會選擇邊一個選項？

1. 撞死行人
2. 撞上隧道牆壁並會將你撞死
3. 其他(請註明)
4. 無答案/拒絕回答

**Q12.** 咁面對「無人駕駛自動汽車系統」嘅道德困境，你認為應該優先考慮邊類人士既意見？係相關公共機構，包括政府機關及法定機構、私人企業、第三方專業人士、受影響嘅市民，定係其他？

1. 相關公共機構(包括政府機關及法定機構)
2. 私人企業
3. 第3方專業人士
4. 受影響嘅市民
5. 其他(請註明)
6. 無答案/拒絕回答

**Q13.** 如果自動駕駛系統需要施加某啲限制，例如禁止市民穿著某類圖案嘅衣服，從而降低計算嘅複雜性並提升系統嘅穩健性。喺系統穩定性，同埋個人自由之間，你認為邊樣價值重要啲呢？你可以俾 1 至 7 分，愈接近 1 分代表個人自由比較重要，愈接近 7 分代表系統穩健性比較重要，中間嘅 4 分代表難以取捨？1 至 7 分，你俾幾多分？

Freedom and autonomy are more important 自由及自主較重要 1			Hard to determine 難以取捨			Beneficence, do no harm and robustness are more important 行善、不傷害及穩健較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

#### 場景 4：機械犬監控

紐約市同新加坡嘅警察部門喺公共地方配置咗機械犬，機械犬會監察市民，並透過內置咪提醒市民喺疫情期間保持社交距離，如果需要對違規行為採取行動，機械犬會進一步通知警方。然而，公眾嘅反應各有不同。紐約市市民傾向反對，新加坡則傾向支持。

**Q14.** 你接唔接受香港政府喺公共地方配置機械犬呢？如果 1 分代表完全不能接受、7 分代表完全可以接受、4 分代表難以決定，1 至 7 分，你會俾幾多分？

Totally not acceptable 1 完全不能接受			Hard to determine 難以決定 4			Totally acceptable 7 完全可以接受
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

頭先嘅情景俾你地了解人工智能嘅應用情況同埋所面對嘅一啲價值取捨。依家想你地就住人工智能應用上，一啲價值同埋原則嘅重要性作出評分。

【Q15-Q19 次序由電話隨機排列顯示】

**Q15.** 喺人工智能應用上，你認為系統透明度有幾重要呢？如果 1 分代表非常不重要、7 分代表完全重要，1 至 7 分，你會俾幾多分？

Least Important 較不重要 1						Most Important 較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

**Q16.** 喺人工智能應用上，你認為系統無偏差有幾重要呢？如果 1 分代表非常不重要、7 分代表完全重要，1 至 7 分，你會俾幾多分？

Least Important 較不重要 1						Most Important 較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

**Q17.** 喺人工智能應用上，你認為保障個人私隱有幾重要呢？如果 1 分代表非常不重要、7 分代表完全重要，1 至 7 分，你會俾幾多分？

Least Important 較不重要 1						Most Important 較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

**Q18.** 喺人工智能應用上，你認為保障個人自由有幾重要呢？如果 1 分代表非常不重要、7 分代表完全重要，1 至 7 分，你會俾幾多分？

Least Important 較不重要 1						Most Important 較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

**Q19.** 喺人工智能應用上，你認為系統穩健性有幾重要呢？如果 1 分代表非常不重要、7 分代表完全重要，1 至 7 分，你會俾幾多分？

Least Important 較不重要 1						Most Important 較重要 7
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1-7 分：\_\_\_\_\_分

8. 無答案/拒絕回答

Trade-offs 取捨

Q20. 如果呢五項價值只能夠保留其中三項，你會保留邊三項呢？1.系統透明度、2.系統無偏差、3.系統穩健性、4.保障個人私隱、5.保障個人自由？請選擇三項，並由認為最重要嗰項開始排序。

- |             |             |             |
|-------------|-------------|-------------|
| (第 1)       | (第 2)       | (第 3)       |
| 1. 系統透明     | 1. 系統透明     | 1. 系統透明     |
| 2. 系統無偏差    | 2. 系統無偏差    | 2. 系統無偏差    |
| 3. 系統穩健性    | 3. 系統穩健性    | 3. 系統穩健性    |
| 4. 個人私隱     | 4. 個人私隱     | 4. 個人私隱     |
| 5. 個人自由     | 5. 個人自由     | 5. 個人自由     |
| 6. 無答案/拒絕回答 | 6. 無答案/拒絕回答 | 6. 無答案/拒絕回答 |

	Preserve first	Preserve as second	Preserve as third
Transparency 透明	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fairness and bias-free 公平無偏差	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Privacy 私隱	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Freedom and autonomy 自由及自主	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Beneficence, do no harm, and robustness 行善、不傷害及穩健性	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q21. 喺接觸到上述人工智能既倫理價值問題之後，有冇改變到你對於人工智能既接受程度？係接受多咗、無改變，定係接受少咗？

1. 接受多咗
2. 無改變
3. 接受少咗
4. 無答案/拒絕回答

## 第三部份 被訪者基本資料

最後係問一啲基本資料，只係用嚟做統計分析。

**DM2. 請問你依家幾多歲呢？**

1. 18-25 歲
2. 26-35 歲
3. 36-45 歲
4. 46-55 歲
5. 56-65 歲
6. Prefer not to say 唔肯講 不願透露

**DM3. 請問你接受教育到乜嘢程度呢？**

1. 小學或以下 Primary school or below
2. 初中 Secondary school
3. 高中
4. 專上非學位 Tertiary, non-degree course
5. 專上學位 Tertiary, degree course
6. 研究院 (碩士或博士學位) Master degree/Doctoral degree
7. 拒答 Refuse to answer

**DM4a. 請問你依家係在職人士、學生、家務料理者、退休人士，定係待業人士呢？**

1. 在職人士
2. 學生 (訪問完結)
3. 家務料理者 (訪問完結)
4. 退休人士 (訪問完結)
5. 待業人士 (訪問完結)
6. 其他非在職人士(例如傷殘人士) (訪問完結)
7. 拒絕回答 (訪問完結)

**DM4b. (只問在職人士) 請問你現時嘅職位係乜嘢呢？**

(不讀出答案)

1. 經理及行政人員 Managers and administrators
2. 專業人員 Professionals
3. 輔助專業人員 Associate professionals
4. 文員 Clerical support workers
5. 服務工作及商店銷售人員 Service and sales workers
6. 工藝及有關人員 Craft and related workers
7. 機台及機器操作員及裝配員 Plant and machine operators and assemblers



8. 非技術工人 Elementary occupations
9. 漁農業熟練工人 Skilled agricultural and fishery workers
10. 拒絕回答 Refuse to answer

**DM5.** (只問在職人士) 請問你現時從事邊個行業呢？

1. 製造業 Manufacturing Industry
2. 建造業 Construction Industry
3. 食肆／酒店 Restaurants / Hotels
4. 運輸 Transportation Industry
5. 倉務 Warehouse Duties
6. 出入口貿易 Import / Export Trade
7. 批發／零售 Wholesale / Retail
8. 銀行及金融 Banks and Finance Sector
9. 保險 Insurance
10. 法律、會計、專業資訊服務 Law, Accountancy, Professional Information Services
11. 商業服務 Commercial Service
12. 房地產 Property
13. 資訊科技 Information Technology (IT)
14. 通訊業 Telecommunication
15. 傳媒 Media
16. 電影／娛樂事業 Film / Entertainment Industry
17. 教育 Education
18. 醫療、衛生及福利 Medical, Hygiene and Welfare Sector
19. 政府／公共事務 Government / Public Affairs
20. 其他個人服務 Other Personal Services
21. 石油及能源 Oil, Energy, Resources and Utilities
22. 其他(請註明) Others (Please specify)
23. 拒答 Refuse to answer

呢次訪問完成啦，多謝你接受訪問。唔該晒，拜拜！