AI Governance for Balanced Development

—

SenseTime Annual Report on AI Ethics & Governance

2022

CONTENTS

[About SenseTime]

SenseTime is a leading AI software company founded in Hong Kong in 2014, focused on creating a better AI-empowered future through innovation. Upholding a vision of advancing the interconnection of physical and digital worlds with AI, driving sustainable productivity growth and seamless interactive experiences, SenseTime is committed to advancing AI research, developing scalable and affordable AI software platforms that benefit businesses, people and society, as well as attract and nurture top talents to shape the future together.

With our roots in the academic world, we invest in original and cutting-edge research that allows us to offer and continuously improve industry-leading, full-stack AI capabilities, covering key fields across perception intelligence, decision intelligence, AI-enabled content generation and AI-enabled content enhancement, as well as key capabilities in AI chips, sensors and computing infrastructure. Our proprietary AI infrastructure, SenseCore, allows us to develop powerful and efficient AI software platforms that are scalable and adaptable for a wide range of applications. Our technologies are trusted by customers and partners in many industry verticals including Smart Business, Smart City, Smart Life and Smart Auto.

SenseTime has been actively involved in the development of national and international industry standards on data security, privacy protection, ethical and sustainable AI, working closely with multiple domestic and multilateral institutions on ethical and sustainable AI development. SenseTime was the only AI company in Asia to have its Code of Ethics for AI Sustainable Development selected by the United Nations as one of the key publication references in the United Nations Resource Guide on AI Strategies published in June 2021.

SenseTime Group Inc. (stock code: 0020.HK) has successfully listed on the Main Board of the Stock Exchange of Hong Kong Limited (HKEX). We have offices in markets including Hong Kong, Mainland China, Macau, Taiwan, Japan, Malaysia, Singapore, South Korea, Saudi Arabia and United Arab Emirates, among others, as well as a presence in Thailand, Indonesia, and the Philippines.

[About the Report]

SenseTime Group Inc. (hereinafter referred to as SenseTime, the Company, or we) takes the initiative in reporting our AI governance progress to the public and invites public engagement in our AI governance process.

Since 2020, SenseTime has been proactively publishing ***the Annual Report on AI Ethics and Governance*** to foster understanding, communication, and interaction between SenseTime, stakeholders, and the public. We aim to promote the development and deployment of "responsible and verifiable AI" through timely disclosure of our AI governance practices.

SenseTime's ***Annual Report on AI Ethics and Governance 2022*** was published globally in September 2022 with both Chinese and English versions. Should there be any suggestions or comments on this report, please contact us at:
aiethics.committee@sensetime.com

SENSETIME I AI FOR A BETTER TOMORROW

[Executive Summary]

Over the years, AI governance has moved from a discussion on principles and policies into technical verification for industry applications. In recent years, regulators and organizations in countries and regions such as Singapore, the European Union, the United States, China have released several AI governance toolkits and sandboxes to promote the implementation of AI governance.

SenseTime is of the view that the technical verification of AI governance is twofold: one is to verify the practicability of principles, guidelines, and policy requirements through practice, while the other is to verify the degree to which AI ethics standards have been implemented by relevant parties through technical or management tools.

We consider AI governance as a dynamic process driven by value, supported by technical tools, implemented with collaboration, and achieved through hierarchical progression. The hierarchy of AI governance comprises four layers: Functional, Reliable, Controllable and Trustworthy, covering the functionality, security and robustness, controllability, and ethical requirements of AI governance.

We believe that AI development and governance should go hand-in-hand, their functions and relationship like that of a nut and bolt, complementing each other and are indispens able. The construction of an AI governance system should not be simply verbal or a quote on paper but should also be traceable and well-documented. Therefore, for the first time in the industry, SenseTime proposes the development of "responsible and verifiable" AI as our vision for AI governance.

To implement AI governance, we have embedded our ethics values in our products through diffusing the principles into “ethics by design” standards and an ethics review matrix. SenseTime is among the first companies that have set up an Ethics and Governance Committee in the industry and made AI governance a strategic priority. The committee is comprised of both internal and external experts and ensures that our business strictly adheres to recognized ethical principles and standards.

In response to ethics risks at the data, algorithm, and application levels, SenseTime has established risk control mechanisms covering the entire life cycle of our products and has begun to see a close-looped AI governance system forming. At the data level, we have established a Privacy Protection Assessment Mechanism, and a Data Security and Personal Information Protection Committee to conduct the assessment covering the acquisition, storage, transmission, and processing of personal information, so as to ensure that our products comply with the design requirements for “privacy protection by default”.

At the algorithm level, we have established an Algorithm Security Assessment Mechanism, and an Algorithm Security Management Working Group. The working group is tasked with classifying and managing algorithms according to their data type, business scenario, ethics risk level, data quality, storage status, application scale, data importance, intervention degree on the user’s behavior and conducting security assessments on algorithm risks arising from technical limitations, algorithm design, software defects, data security, framework security, and other related dimensions.

At the application level, we have established an Ethics Risk Classification Management Mechanism and an Ethics Risk Review Team to carry out graded and targeted ethics risk management throughout the entire product life cycle of design, development, deployment, and operation. We have also set up supporting processes for self-inspection, assessment and review of risks, and follow-up reviews. We classify ethics risks from low to high, spanning five levels from E0 to E4, based on the impact of final product safety, personal rights and interests, market fairness, public safety, and environmental health.

To promote the development of responsible and verifiable AI, we have developed a series of internal management tools and technical tools covering data governance, algorithm evaluation, model examination, and ethics review. To cultivate an organizational culture for AI governance, we have published Ethics and Governance Policy, SenseTime AI Ethics and Governance Committee Management Charter, and Guidelines for Ethics risk Review, to provide a set of guidelines and clearly-defined standards for aligning employee’ understanding and actions on AI ethics and governance. To better educate and communicate with employees regarding AI governance, we have established regular training and community engagement platforms.

So far, our AI ethics and governance system and related technical tools have been recognized by various third-party organizations, such as the Harvard Business Review and Artificial Intelligence Industry Alliance (AIIA). Our first White Paper on AI Sustainable Development was also included in the United Nations’ Resource Guide on Artificial Intelligence Strategies.

SENSETIME I AI FOR A BETTER TOMORROW

I.

Overview of AI Development and Governance

> AI has entered a new stage of development since 2010 where computing power and data became the main driving force. At this stage, AI is no longer based on just human cognition to some extent. In fact, its rules have been far beyond current human cognition, sparking a wave of discussions on its governance.
>
> **- Dr. Xu Li, Executive Chairman of the Board and CEO of SenseTime**

Over the past decade, driven by deep learning, big data, and Moore's Law, AI has made many remarkable breakthroughs, and has been commercialized in various segments such as computer vision, natural language processing, and speech recognition. Today, AI is being widely deployed in city management, as well as in industries such as education, finance, medical care, retail, transportation, entertainment, and manufacturing, amongst others, and is expanding into other fields of knowledge exploration such as scientific research and the Arts. AI is increasingly being recognized and adopted as a general-purpose technology, which accelerates the arrival of the era of ubiquitous intelligence.

Throughout history, general-purpose technologies have inevitably brought fundamental changes to the existing structure of society, while revolutionizing social productivity. This law applies to AI too. In particular, when data-driven AI breaks the boundaries of human cognition, its impact on the existing social structure will be even greater. Even at the present stage of "weak (narrow) AI", concerns about the robustness and fairness of automated decision-making systems and the abuse of recommender systems, deep synthesis, and biometric information identification technologies clearly indicate that if the industry and all stakeholders don't act proactively to seek broad consensus on how AI should be designed, developed, and deployed, the journey to the era of ubiquitous intelligence will increasingly face trust challenges.

For this reason, AI governance has garnered attention from enterprises, government agencies, international organizations, social groups, and other stakeholders in the past decade and made remarkable progress.

Looking at the global AI governance process, it has so far gone through three stages of development, and has now entered the stage of implementation:

Figure 1 Global AI Governance Development History

Source: Institute for AI Industry Research, SenseTime

- **Stage 1.0 of AI Governance began in 2016 with a focus on principle discussion.**

  In their study, Jessica Fjeld et al. from Harvard University identified September 2016 as the beginning of AI Governance 1.0 with the publication of the ***Principles of Partnership on AI*** by a group of tech giants that included Google, Facebook, IBM, Amazon, and Microsoft. [1]After analyzing 84 principles or guidelines regarding AI ethics worldwide, Anna Jobin et al. also found that 88% of them had been published after 2016, and the number of documents released by private enterprises and government agencies accounted for 22.6% and 21.4%, respectively.[2]

- **Stage 2.0 of AI Governance began in 2020 with a focus on policy discussion.**

  In February 2020, the European Commission released the ***White Paper on Artificial Intelligence***[3], the first in the world to propose a risk-based regulatory framework for AI governance. Since then, major countries have followed suit and explored regulations on AI-related technologies and applications to varying degrees. For this reason, 2020 is often referred to as the starting point of AI regulation. According to OECD statistics, more than 700 AI policy initiatives have been proposed by over 60 countries around the world, while Deloitte Global predicts that 2022 will see even more discussion on the systematic regulation of AI.[4][5]

1 https://cyber.harvard.edu/publication/2020/principled-ai
2 https://doi.org/10.1038/s42256-019-0088-2
3 https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
4 https://www.pymnts.com/news/regulation/2022/oecds-principles-can-guide-governments-to-design-ai-regulatory-frameworks/
5 https://www2.deloitte.com/global/en/insights/industry/technology/technology-media-and-telecom-predictions/2022/ai-regulation-trends.html

- **Stage 3.0 of AI Governance began in 2022 with a focus on technical verification.**

  In 2022, there are an increasing number of initiatives aimed at verifying how AI governance is implemented, as the global AI governance process continues to move forward and concepts such as trustworthy and responsible AI gain greater traction. Within the public sectors, the government of Singapore launched the world's first open-source testing toolbox for AI governance, "AI.Verify", in May 2022. In June 2022, the Spanish government and the European Commission introduced the first pilot project for an AI regulatory sandbox. On the market side, the Responsible Artificial Intelligence Institute, a US-based AI governance research institute has released a Responsible AI Certification Program to provide responsible AI certification services to enterprises, organizations, and institutions.

At stage 3.0, we are of the view that the technical verification of AI governance is twofold: one is to verify the practicability of principles, guidelines, and policy requirements through practice, while the other is to verify the degree to which AI ethics standards have been implemented by relevant parties through technical or management tools. Next, implementing AI governance needs to properly address challenges around the following three relationship groups:

Principle <> Action	Policy <> Practice	Technology <> User
An IBM study found that while more than half of surveyed organizations have published or publicly endorsed common principles of AI ethics, less than a quarter have actually put them into practice.	In the process of promoting AI governance, policy requirements need to be transformed into practical standards that can be implemented by technical and business personnel	An IBM study found that only 40% of surveyed consumers believed businesses were responsibly and ethically developing and deploying new technologies, and users lacked adequate understanding on AI governance.

Figure 2: Relationship Need to Be Properly Addressed in Stage 3.0 of AI Governance

Source: Institute for AI Industry Research, SenseTime

- **The relationship between principles and action.**

  An IBM study found that while more than half of surveyed organizations have published or publicly endorsed common principles of AI ethics, less than a quarter have actually put them into practice. [6]AI ethics and governance still face many challenges in practice: First, the integration of AI governance into the existing organizational structure. AI governance involves information security, data governance and other overlapping areas with the existing organizational structure. Issues such as overlapping responsibilities and unclear scope of work have led to certain constraints on the promotion of implementation at the organizational level. Second, AI governance has not been truly integrated into the industry's business value chain. When promoting AI ethics and governance, the lack of clear returns on investment may lead to organizations paying less attention to AI ethics and governance and lagging in implementation. Third, there is a lack of consensual standards on how to conduct AI governance.

- **The relationship between policy and practice.**

  Policymakers and technology developers share different perspectives, positions, and understandings of policy implications. In the process of promoting AI governance, policy requirements need to be transformed into practical standards that can be implemented by technical and business personnel. The following four aspects needs to be kept in mind when promoting AI governance: First, policy formulation needs to take into account the dynamics and diversity of industries and AI applications, to foster a benign environment conducive to industry development. Second, different institutions, public agencies, and countries need to strive to promote the interoperability of AI governance standards. Third, AI governance practitioners need to be involved in policy formulation processes, so as to make policies more practical. Fourth, policy makers and industry practitioners need to seek consensus in defining issues related to AI governance.

- **The relationship between technology and users.**

  At present, AI governance remains within the purview of professional discussions and corporate governance, with end-users yet to be included in the loop of AI governance. Consequently, there are often misunderstandings on AI governance issues within the market and society. For example, some users may define temporary technical issues as long-term governance challenges. Research by IBM found that only 40% of surveyed consumers believed businesses were responsibly and ethically developing and deploying new technologies, and users lacked adequate understanding on AI governance.[7]Therefore, the relationship between technology and users should be properly addressed when promoting AI governance. Technology providers need to seek to explain technology from the users' perspectives, while AI governance institutions need to clarify the real challenges facing governance, deepen users' understanding on AI governance, and give users the opportunity to participate in AI governance. In addition, it is necessary for enterprises to increase investment in AI governance-related technical tools, to improve the verifiability of AI governance.

6 https://www.ibm.com/downloads/cas/VQ9ZGKAE

7 https://www.ibm.com/downloads/cas/VQ9ZGKAE

II.

How We Think of AI Governance

> Technology and human activities today are deeply integrated. We need ethics and humanism to promote the healthy development of science and technology that is beneficial to mankind. Therefore, we attach great importance to AI ethics and governance.
>
> **– Dr. Xu Li, Executive Chairman of the Board and CEO of SenseTime**

SenseTime has long attached great importance to AI governance. Since 2019, it has established and maintained a Global AI Ethics Risk Registry internally, which covers hundreds of AI ethics best practices and alerts. While closely tracking the development of global AI governance, we strive to deepen our knowledge and understanding of AI governance as we seek more systematic thinking on AI governance issues.

Based on our in-depth analysis of global risk cases, **we observed that the governance challenges inherent to the AI era stem mainly from three levels: data, algorithm, and application. Specifically:**

- **At the data level,** the risk primarily involves privacy protection, data governance, and data quality. Of these, privacy protection risk refers to issues of privacy violation during AI development, testing, and operation, and it is one of the major problems to be addressed in current AI applications. Data quality risk refers to flaws that may exist in training data sets and field data collected for AI, as well as the corresponding adverse effects. This is also a type of data risk specific to AI. Data security risk refers to the security protection of data held by enterprises in the process of AI development and application, which involves the entire life cycle of data, i.e. collection, transmission, storage, processing, and circulation.

- **At the algorithm level,** the risk mainly involves algorithm decision-making, black box algorithm, and algorithm security. Of these, algorithm decision-making risk refers to the inability to predict the reasons and effects of decisions made by AI systems, due to the unpredictability of the results of algorithmic reasoning and the cognitive limitations of human beings. For example, the problem of liability fixation is a typical one. Black box algorithm risk refers primarily to interpretability risk resulting from opaque decision-making and the inability to be fully explained due to the complexity of neural network algorithms. Algorithm security risk refers to the risk caused by the leakage or malicious modification of model parameters and insufficient fault tolerance and elasticity.

- **At the application level,** the risk involves algorithmic bias, ethical conflict, and labor substitution among others. For example, due to subjective factors, or bias contained in training data and data input during self-learning process, machine-learning algorithms may introduce bias into its decision-making process. The risk of algorithm abuse can result from ill induction to users and excessive application of algorithms. AI can also impact employment negatively in the long term, exacerbating unfair competition and market dominance and causing dilemmas and risks in defining responsibility.

Based on our own experience and observation of the global AI governance process, we consider AI governance as a dynamic process driven by value, supported by technical tools, implemented with collaboration, and achieved through hierarchical progression. The hierarchy of AI governance comprises four layers: Functional, Reliable, Controllable and Trustworthy, covering the functionality, security and robustness, controllability and ethical requirements of AI governance.

- Functional means that an AI system can satisfy the application requirements in terms of function and performance.

- Reliable means that an AI system can satisfy the requirements of the deployment environment and sustainable operation in terms of security and robustness.

- Controllable means that an AI system can adequately protect the independent will and rights of human beings and guarantee a human's control over the system on the functional level.

- Trustworthy means that the design and application of an AI system that conforms to human values.

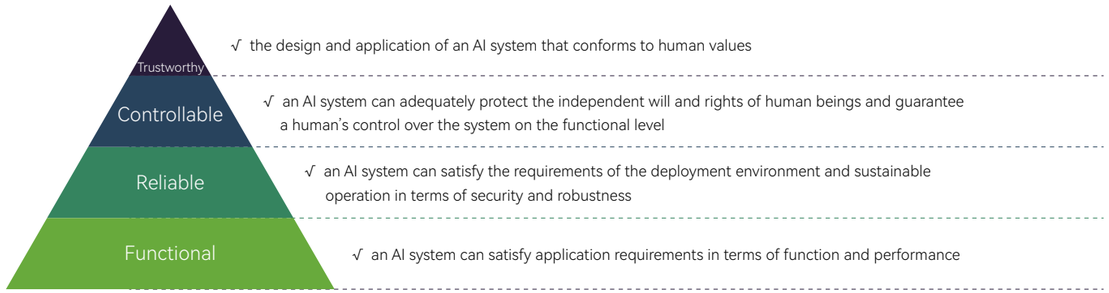


Figure 3: Hierarchy of AI Governance

Source: SenseTime

III.

AI Ethics for "Balanced Development"

> AI development and governance should go hand-in-hand. If governance is applied prematurely, it may limit AI development. However, if it lags behind, the consequences could be catastrophic and the cost of reparative governance will be high.
>
> **– Dr. Xu Li, Executive Chairman of the Board and CEO of SenseTime**

Given the state of its development and commercialization, AI technology and its applications are still in the early stages, and AI-related economic form, industry ecosystem, and business model are still in the exploratory stages. Like other general-purpose technologies that have emerged throughout the course of history, the healthy and sustainable development of AI technology and relevant industries not only requires an innovative space that is consistent with the current state of its development, but also needs appropriate guidance and guardrails. **Hence, we believe that AI development and governance should go hand-in-hand, their functions and relationships are like that of a nut and bolt, complementing each other and are indispensable.**

Based on our understanding of AI development and governance, we unveiled the "AI Ethics for Balanced Development" Report in 2021, to further crystalize SenseTime's three core ethical principles of responsible AI: sustainability, human-centric approach, and controllable technology. Specifically, the concept of "balanced development" advocates AI development should be complemented with governance and promotes the healthy and sustainable development of the AI industry, as well as the digital transformation of the overall economy and society through AI governance.

- "Human-centric" advocates respecting, accommodating, and balancing differences in historical, cultural, social, and economic development among different countries and regions, and pursuing consensus among different cultures. Meanwhile, we should also ensure the protection of human rights and privacy and deploy technology without prejudice.

- "Controllable technology" advocates that AI is developed by and for humans and therefore, should be controlled by humans. Correspondingly, its controllers, i.e., humans, should be responsible for its actions.

- "Sustainability" advocates the sustainable development of society, economy, culture, and the environment, and promotes openness and inclusive innovation.

To embed the values in our products, **we have further diffused the principles into "ethics by design" standards. These high-level standards are as follows:**

- Respect for human rights. Human freedom and dignity must be protected along with other basic rights, as globally recognized ethics standards are upheld, human development and life experience can be enhanced without harming human status.

- For good. Sustainable development for humans must be protected, just as the interests of vulnerable groups are protected. Adhere to the basic concepts of human ethics and morality, and the applications should be reasonable, legal, and compliant.

- Free from bias. The data used should overall be objective, neutral, and representative and balance universal applicability with the needs of special populations.

- Protection of privacy. The collection and usage of personal information should adhere to the principle of data minimization. In particular, the processing of sensitive personal information must obtain the consent of the information subject or specific circumstances prescribed by law.

- Reliable and controllable. Within a certain period of time, under certain conditions, specific functions can be implemented without failure. Even in the case of failure, effective shutdown and human takeovers can be implemented.

- Transparent and explainable. Priority should be given to algorithmic models that can be explained. Users should be provided with clear, understandable, and satisfactory descriptions of the product operating mechanism, and the limitations and potential risks of the product should be presented clearly.

- Verifiable. It should be possible to verify the algorithmic model and its results repeatedly under the same or similar conditions.

- Accountable. The rights and obligations of the subjects in R&D, design, manufacturing, operation, and service should be clearly defined, and relevant mechanisms should be available to trace the models and data behind the output results.

SENSETIME I AI FOR A BETTER TOMORROW

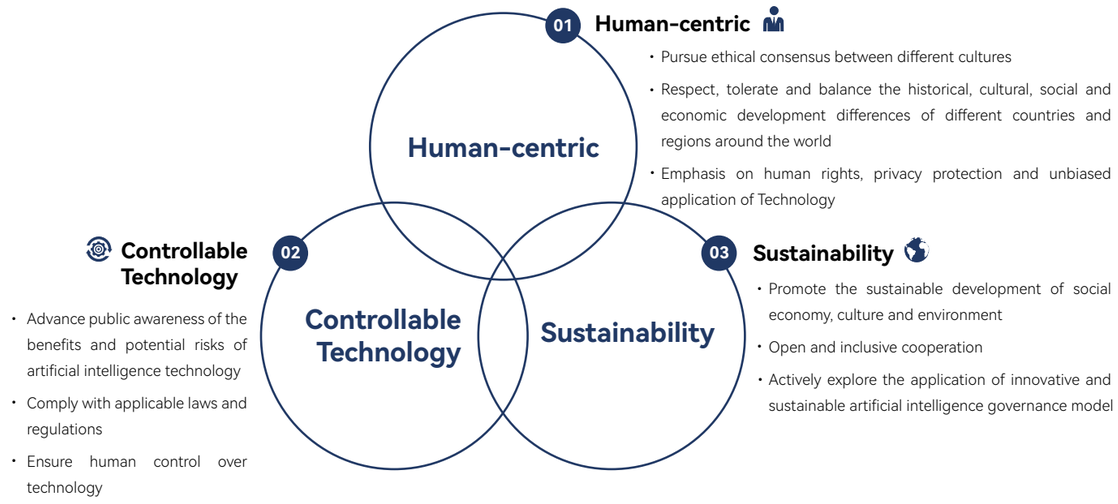


Figure 4: Three Core Ethical Principles of Responsible AI

Source: SenseTime

IV.

Responsible and Verifiable AI

> With the rise of new technologies, the ethics of science and technology faces many new topics that require joint research involving every part of society. As an industry leader, SenseTime has the responsibility to adhere to high standards of AI ethics.
>
> **– Zhang Wang, Vice President of SenseTime, Chairman of the AI Ethics and Governance Committee.**
>
> The boundaries and core requirement of managing AI ethics risks should be the development of responsible AI. In terms of governance and compliance, enterprises should think and act ahead.
>
> **– Yang Fan, Co-founder and Vice President of SenseTime, Member of the AI Ethics and Governance Committee.**

Trust must be the core of the business community and the basic premise for the widespread acceptance of emerging technologies. SenseTime, as an innovative scientific enterprise in the field of AI, has always regarded the trust of the market and users as the key to its development. Since SenseTime was founded, all our actions have been guided by responsible AI development and deployment. At the same time, we believe that responsible AI is not only a principle, but also concrete and implementable. The key to achieving this goal is to build a holistic AI governance system.

After exploring AI governance in-depth, we have realized that the construction of an AI governance system should not be simply verbal or a quote on paper, but should also be traceable and well-documented. Therefore, **for the first time in the industry, we propose the development of "responsible and verifiable" AI as our vision for AI governance. Specifically, "responsible and verifiable" AI should meet the following basic requirements:**

- Responsible for people. AI systems should respect and protect the dignity and rights of people and contribute to the health and well-being of people.
- Responsible for society. AI systems should respect and adapt to the customs and habits of different cultures and contribute to the healthy and sustainable development of society.
- Responsible for the environment. The development of AI systems should be mindful of its environmental impact, and its use should be beneficial to the sustainable development of the environment.
- Accountability for responsible parties. Responsible parties should be clearly defined for the entire life cycle and all modules of every AI system, so as to ensure accountability.
- Risks evaluated. AI systems should go through adequate ethics risk assessments before going online.
- Verifiable governance process. The entire life cycle of AI systems and relevant governance processes should have complete technical logs and documentation.
- Verifiable governance results. The implementation of AI governance standards over the entire life cycle of AI systems should be supported by technical and administrative tools.

Figure 5: Core Requirements for Responsible and Verifiable AI

Source: SenseTime

V.
How We Implement AI Governance

(1) Organization Innovation

SenseTime is one of the first companies in the industry to establish an AI Ethics and Governance Committee and make AI governance a strategic priority.

To systematically respond to the ethics risks of AI at different levels, such as data, algorithms, and applications, SenseTime officially established the AI Ethics and Governance Committee in January 2020 to develop an AI ethics governance system. The AI Ethics and Governance Committee is comprised of two external members and four internal members, who come from technical, engineering, legal, ethics, and related professional backgrounds. The Committee is also supported by a secretariat, an expert advisory group, and an executive working group to ensure the independence, transparency, professionalism, and effectiveness of ethics governance. In addition, to ensure the efficient operations of the AI Ethics and Governance Committee and to strengthen the compliance of ethics standards by all employees, SenseTime has published the ***AI Ethics and Governance Committee Management Charter, Ethics and Governance Policy, Guidelines for Ethics risk Review*** and other ethics-related corporate policies.

Daily Work Responsibilities

- Formulation of ethical strategic development plans.
- Product ethics review and risk control.
- General ethical knowledge training and promotion.
- Ecological construction of ethical governance.
- Participation in discussion and formulation of ethical standards.
- Extensive joint research.

Important Work Content

Independent opinions on major issues of the company and AI Ethics Committee (see below for details).

- Decision-making on major issues of corporate ethical strategy.
- Setting and adjusting development plan of AI Ethics Committee.
- Nominating, appointing, and dismissing expert consultants of AI Ethics Committee.
- Other matters stipulated by relevant laws, administrative regulations, departmental rules, normative documents, and articles of association.

Figure 6: Responsibilities of the AI Ethics and Governance Committee

Source: SenseTime

At the same time, to strengthen AI governance at all levels, we have systematically enhanced the coordination among internal organization structures and workflows and provided the Information Security Management Committee with the ability to conduct privacy protection assessment and algorithm security assessment.

(2) Mechanism Innovation

In response to ethics risks at the data, algorithm, and application levels, SenseTime has established risk control mechanisms covering the entire life cycle of our products, and begun to see a close-looped AI governance system forming.

At the data level, we have established a Privacy Protection Assessment Mechanism, and a Data Security and Personal Information Protection Committee to conduct the assessment covering the acquisition, storage, transmission, and processing of personal information, so as to ensure that our products comply with the design requirements for "privacy protection by default".

At the algorithm level, we have established an Algorithm Security Assessment Mechanism, and an Algorithm Security Management Working Group. The task of the working group is to classify and manage the algorithms according to their data type, business scenario, ethics risk level, data quality, storage status, application scale, data importance, intervention degree on the user's behavior, and to conduct security assessments on algorithm risks arising from technical limitations, algorithm design, software defects, data security, framework security, and other related dimensions.

At the application level, we have established an Ethics Risk Classification Management Mechanism and an Ethics Risk Review Team to carry out graded and targeted ethics risk management throughout the entire product life cycle of design, development, deployment, and operation. We have also set up supporting processes for self-inspection, assessment and review of risks, and follow-up reviews. We classify ethics risks from low to high, spanning five levels from E0 to E4, based on the impact of final product safety, personal rights and interests, market fairness, public safety, and environmental health:

- E4 products: prohibited products. These refer to AI products that deviate from SenseTime's ethics principles and violate the requirements of laws and regulations.
- E3 products: high-risk products. These refer to products directly related to the final product's safety, personal rights and interests, market fairness, public safety, and environmental health.
- E2 products: medium-risk products. These refer to products that have indirect or potentially high impact on the final product's safety, personal rights and interests, market fairness, public safety, and environmental health.
- E1 products: low-risk products. These refer to products that have no obvious impact on the final product's safety, personal rights and interests, market fairness, public safety, and ecological security.
- E0 products: risk-free products. These refer to products that exclude machine learning algorithms and AI functions.

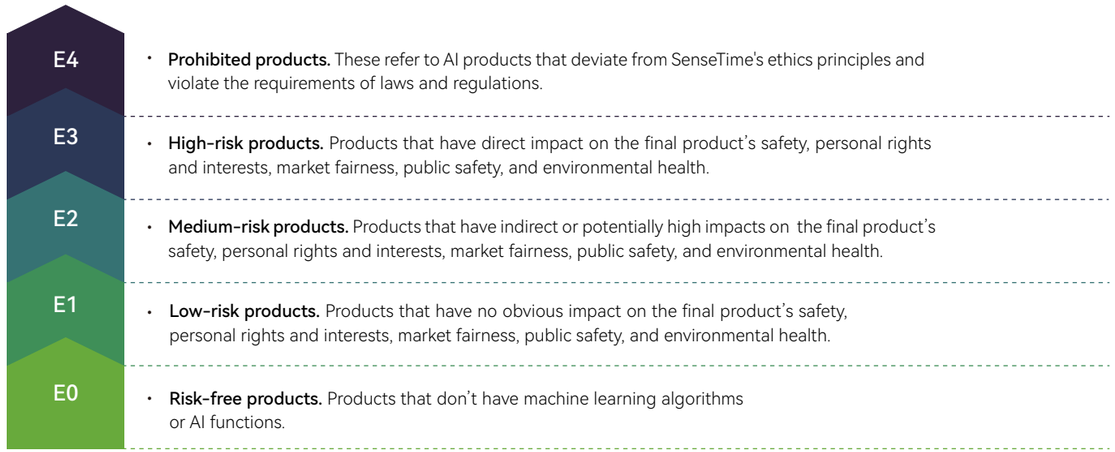


Figure 7: SenseTime's Standard for Ethics Risk Classification

Source: SenseTime

In addition, to ensure the effective implementation of AI governance systems and to promote an ethic-respecting culture in a corporate setting, we have also established Ethics Risk Management Goal-setting Mechanism, Ethics Incidents Reporting and Mitigation Mechanism, and Ethics Governance Quality Control Mechanism.

(3) Tools Development

To promote the development of responsible and verifiable AI, we have developed a series of internal management tools and technical tools covering data governance, algorithm evaluation, model examination, and ethics review.

At the data level, we have developed a unified data governance platform and standardized data collection processes to ensure accuracy, balance, and rationality. Besides, with a data and privacy protection platform, we realized the privacy encryption of data, thus ensuring complete data availability, reliability, and security. At the same time, we have designed a set of personal information protection assessment checklists for the whole process of product development, promoted functional product design oriented to personal information protection, ensured the design process of AI products, and limited the collection and processing (including usage, disclosure, retention, transmission, and disposal) to clearly defined and necessary purposes.

In addition, in the process of data processing, by developing and deploying automatic labeling tools, we reduce the amount of data contacted manually and the risk of introducing human bias at the source of model training. Moreover, the data labeling platform has access control and authentication functions and can only be accessed by certified data labeling personnel.

At present, we have received a number of internationally recognized certifications for network and data security, including The Information Security Management System Certificate (ISO/IEC 27001:2013), The Privacy Information Management System (PIMS) Certificate (ISO/IEC 27701:2019), The Code of Practice for Personally Identifiable Information Protection (ISO/IEC 29151:2017), and The Personal Information Security Management System Certificate (BS10012). Products sold have also obtained Level 3 Certification for Important Information System Classified Protection, and The Trusted Face Certification Special Test Certificate among others.

At the algorithm level, the black box algorithm poses a significant risk of discrediting the algorithm and hindering its explanation. When designing a model, we output various types of information in the code to enable fast traceability in case of an algorithm decision error.

At the same time, by establishing a model inspection platform to test the model for inference attacks and reverse attacks, we can check and score the test factors of the algorithm model against digital world white-box confrontation, digital world black-box query attack robust accuracy, digital world migration attack robust accuracy, physical world sample attack success rate, and model backdoor attack success rate among others to determine whether the algorithm model meets the design requirements. When the algorithm model does not meet the design requirements, we enable the corresponding algorithm repair module to improve security. At the same time, we build an AI firewall at the system level to defend against attacks from adversarial examples. When the model is released, tests are done on the test set defined by the product, and manual testing is conducted to ensure that the scale of the test set is large enough and can reach the required accuracy level.

In addition, based on the algorithm verification and evaluation of data sets in real scenarios, we have developed an algorithm evaluation tool. It is capable of fully evaluating algorithms through full coverage of the algorithm evaluation in main scenarios and long-tail scenarios, diversified evaluation items and rich index systems, as well as sufficient data sets and comprehensive evaluation schemes, for credibility and control of all commercial algorithms.

At the application level, we have designed a set of ethics risk self-examination tools and a review platform in conjunction with the different stages of review. At present, all SenseTime's AI products must undergo ethics risk reviews through the review platform at different stages from project approval and release to online operation. Before the review process, self-inspection tools can be used to prepare for the review. During the review process, we may choose to reject new product proposals, suspend the ongoing product development projects, or withdraw existing products that do not meet our principles and standards.

Due to the above-mentioned practices, our AI ethics and governance system and related technical tools have been recognized by various third-party organizations, such as the Harvard Business Review and Artificial Intelligence Industry Alliance (AIIA). Our first ***White Paper on AI Sustainable Development*** was also included in the United Nations' ***Resource Guide on Artificial Intelligence Strategies.***

Harvard Business Review, Ram Charan Management Practice Award

Figure 8: External Recognition of SenseTime's AI Governance Practices

Source: SenseTime

(4) Fostering Culture

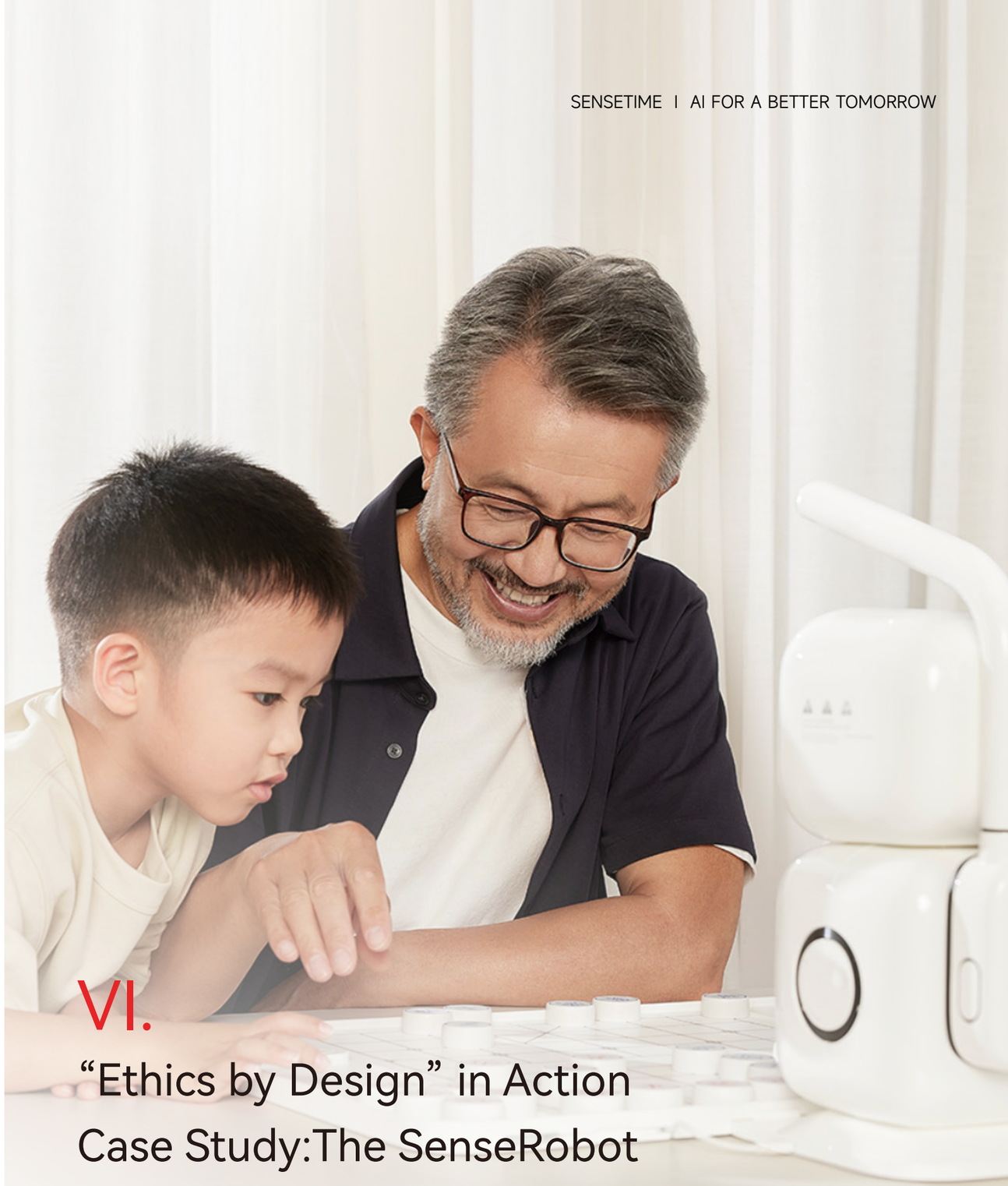
We realize that the key to developing responsible and verifiable AI is to cultivate an organizational culture for AI governance.

With the release of ***Ethics and Governance Policy, SenseTime AI Ethics and Governance Committee Management Charter,*** and ***Guidelines for Ethics Risk Review,*** we provide a set of guidelines and clearly-defined standards for aligning employee' understanding and actions on AI ethics and governance. At the same time, to better acquaint employees with AI governance topics, we have introduced regular trainings and community engagement platforms. We send important trends related to AI governance to all employees weekly and regularly organize seminars, inviting internal and external experts to provide training on AI ethics and governance.

(5) Developing the Ecosystem

We actively participate in standard-setting bodies related to AI ethics and governance such as the National Information Security Standardization Technical Committee and the Institute of Electrical and Electronics Engineers (IEEE) and serve as chair or vice chair in several working groups. At the same time, we have established research cooperation on AI ethics and governance with well-known universities at home and abroad, and research institutes such as Tsinghua University, Shanghai Jiao Tong University, and Artificial Intelligence International Institute.

We jointly launched the Tech4SDG alliance with industry partners to promote the steady development of responsible and verifiable AI. Currently, the Tech4SDG alliance covers 9 countries and regions in Asia, with more than 40 members, including many well-known universities and think tanks from mainland China, Hong Kong, Macau, Singapore, India, and Saudi Arabia, among others.



VI.

“Ethics by Design” in Action

Case Study:The SenseRobot

— An AI Chinese Chess Robot

As mobile phones and tablets become increasingly dominant in our lives, many parents hope to see their children spend less time on their electronic devices. On August 9th, 2022, SenseTime officially launched its first household consumer AI product, SenseRobot, the AI Chinese chess robot. Integrated with SenseTime's leading AI technology and mechanical arm technology, the SenseRobot is a physical robot that can be placed at homes and on the table. Children interacting with SenseRobot need not look at its electronic screen, and are able to learn and play Chinese chess without straining their vision.

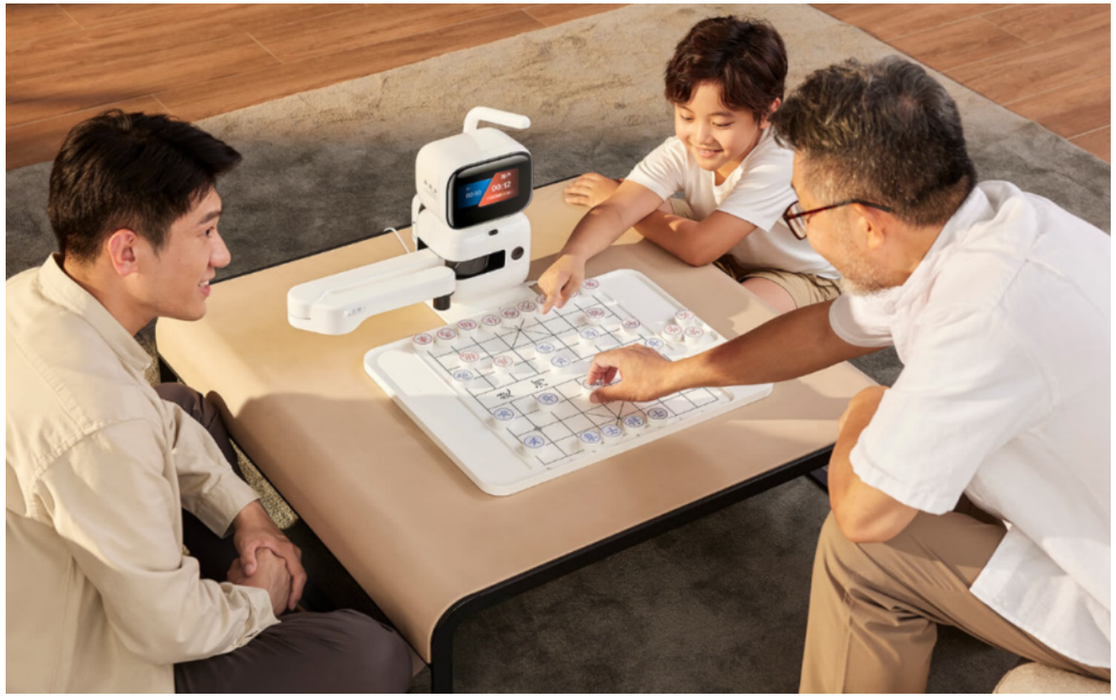


Figure 9: SenseRobot Application Scenario

Source: SenseTime

SenseRobot has a simple and sleek appearance. It presents itself as a little "astronaut", who teaches and plays Chinese chess with children "face-to-face".

Dr. Xu Li, Executive Chairman of the Board and CEO of SenseTime, said at the product launch event, "Our goal is to create a robot that can physically 'think' and 'act' with our leading AI technology, bring industrial grade AI technology into every family, and make real interactions with children and elderly. It can not only accompany the whole development period of children, but also make high technology intuitive, understandable and interesting for elderly. It will bridge the digital divide and build emotional connection with technology, while bringing overall enjoyment to the whole family."

SenseRobot features AI chess learning and various levels of challenges, among others. It can introduce and explain Chinese chess culture, rules and skills of each chess piece to children without previous experience. While training the children, it can also improve their cultural literacy. In addition, it also contains more than 100 endgames and 26 levels of chess competition, so that users not only experience playing with actual Chinese chess pieces, but also enjoy the mental stimulation at various difficulty levels.

With SenseTime's leading AI technology, the SenseRobot has remarkable coordination and can achieve millimeter-level of operation accuracy to ensure the game runs smoothly. Furthermore, it has been certified and authorized by the Chess and Card Management Center of the General Administration of Sport and the Chinese Xiangqi Association, so users can be assessed for levels 16 – 13 in the official Chinese chess level examinations.

Based on feedback from users of the first batch of trials, they said that SenseRobot is a product that brings the family together through fun and games. Through AI deep learning and self-training, SenseRobot's capabilities are at expert-level, and there are suitable activities for both beginners and experienced players. In addition, it also brings the whole family together to come up with solutions for the chess challenges, thereby strengthening the bonding between children, parents, and grandparents. SenseRobot's aim to stimulate minds and bring the family together is an example of human-centric design.

SENSETIME I AI FOR A BETTER TOMORROW

Editorial Committee

Author

Hu Zhengkun

SenseTime
Director of AI Ethics and Governance Research
huzhengkun@sensetime.com

Tian Feng

SenseTime
Dean of Institute for AI Industry Research
tianfeng@sensetime.com

Acknowledgments

Mei Ying，Yan Xintong，Li Yuelu，Liu Zhiyi，Gong Chao，Wu Jingyu，Cheng Jin，Gong Liuting
Qi Weiliang，Liang Ding，Wu Yichao，Wang Yifei，He Conghui，Xin Yuchen

For more information, please follow SenseTime on:

Official Website: **https://www.sensetime.com/**

LinkedIn: **https://www.linkedin.com/company/sensetime-group-limited/**

Facebook: **https://www.facebook.com/sensetimegroup/**