International Journal of Law and Society
2025, Vol. 8, No. 4, pp. 261-290
https://doi.org/10.11648/j.ijls.20250804.11

Research Article

Generative Artificial Intelligence: Research on the Risk of Science and Technology Ethics, Regulatory Mechanism and Governance Strategy

Chunlei Sun*

School of Law, Beijing Institute of Technology, Beijing, China

Abstract

Digital information technologies represented by General Artificial Intelligence (GAI) are becoming the core driving force behind profound transformations in the digital judicial field. GAI may trigger ethical risks across cognitive, subject, knowledge, and data dimensions, including dissemination of biases and discrimination, privacy breaches, fabrication of false information, manipulation of cognition, and even research ethics violations. The root causes of technological ethics issues stem from insufficient stakeholder accountability, imbalance between instrumental and value rationality, self-limitation of digital judicial technologies, and inadequate AI regulatory mechanisms. It is essential to follow practical mechanisms involving technology mediation, digital benevolence, legal supervision, and ethical standard formulation. This approach aims to enhance GAIs role in reinforcing the value concept of "promoting human welfare," improving ethical governance, increasing technical explainability, strengthening corporate self-discipline, and establishing global collaborative mechanisms for comprehensive risk management. GAI undergoes rapid innovation iterations, it continues to empower human society across multiple domains: serving as an essential driving force for innovation and efficiency enhancement in the digital age, while simultaneously raising public ethical concerns due to its inherent uncertainties, risks, and complex disruptive social impacts. Assessing and addressing GAIs ethical risks proves crucial for protecting individual rights, upholding fairness and justice, and promoting social harmony. Therefore, we should approach GAI technological progress with a cautious yet forward-looking ethical perspective, ensuring that GAI truly creates benefits for humanity.

Keywords

Generative Artificial Intelligence, Ethical Risks, Technology for Good, Governance of Technology Ethics, GAI Enhancing Human Well-Being

1. The Proposal

The future has arrived. The academic community has established a relatively unified concept and classification of artificial intelligence. The primary goal of AI research is to enable machines to achieve human-like intelligence, meaning they can handle complex tasks that previously required only human intelligence. Simultaneously, it explores how to develop intelligent artificial systems capable of processing and executing tasks like humans. The development status of AI

*Corresponding author: 18661901918@163.com (Chunlei Sun), whc0313@sina.com (Chunlei Sun)

Received: 3 October 2025; **Accepted:** 13 October 2025; **Published:** 30 October 2025

Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

varies in strength: AlphaGo and Deep Blue fundamentally belong to computational tools, operating through algorithmic programming and mechanical task execution, thus qualifying as weak AI. These are typically domain-specific intelligent systems designed to perform single tasks. In contrast, academia defines strong AI as intelligent machines with perception and self-awareness, possessing genuine reasoning and problem-solving capabilities that approach or even surpass human intelligence levels. There are two types: humanoid AI, where intelligent agents can reason or think like humans; and non-humanoid AI, where intelligent agents can generate consciousness or perception distinct from human cognition. This cognitive approach differs from human reasoning, heralding a new civilization. The "strong AI perspective" holds that computers are not merely tools for studying humans... Its a tool for thinking; conversely, the computer itself is thinking as long as the right program is running." [1].

At present, artificial intelligence is mainly divided into two models: discriminant and generative.

Predictive output under given input conditions typically focuses on category boundaries and decision surfaces, enhancing classification accuracy and demonstrating superior performance in specific tasks. In contrast, generative models concentrate on simulating data generation processes. These models not only perform classification or prediction but also generate new data samples, including various types of data such as text, images, audio, and video. [2]

In other words, generative AI can recognize patterns in training data, learn to predict, categorize, offer personalized suggestions or help make decisions, and quickly create new content based on user prompts. [3]

Currently, major countries worldwide have made significant progress in generative AI innovation. China has listed artificial intelligence as a national strategy and has become one of the key forces in global AI innovation, achieving major breakthroughs in fields such as natural language processing, computer vision, and intelligent robotics, continuously driving the innovative development of AI technologies. Large language model products like Baidus "Wenxin Yiyang" and iFLYTEKs "Xinghuo" are rapidly iterating. [4]

DeepSeek has garnered significant global attention. Generative AI, recognized as a powerful creative engine with immense potential, has greatly fueled humanitys aspirations for Artificial General Intelligence (AGI), marking a transformative leap in AI innovation. As the latest stage of AI development, this technology "can reshape existing workflows, revolutionizing daily life and production methods across content creation, autonomous driving, retail, healthcare, and other sectors while unlocking vast market opportunities." [5]

As stated in the book The Age of Artificial Intelligence and the Future of Humanity, unlike previous technologies, artificial intelligence is likely to change the trajectory of society and the course of history, constantly integrate into our lives and bring about a new world. [6]

General Artificial Intelligence (AGI) represents the broader scope of AI development, while narrow AI is commonly referred to as traditional artificial intelligence. This concept embodies the foundational assumptions and ultimate goals of current AI progress, often featured in science fiction films but not yet achievable in reality. The term was first introduced by American researchers Ben Goertzel and Cassio Pennachin in 2007. AGI refers to self-learning intelligent systems capable of autonomous development, representing a higher evolutionary stage of AI that can learn, improve, adapt, and solve problems independently without human intervention. "Unlike conventional AI designed for specific functions, AGI aims to simulate human cognitive capabilities. The defining characteristic of human intelligence lies in its ability to make adaptive adjustments to different environments and handle diverse tasks—even previously unexplored ones." [7, 8]

Generative Artificial Intelligence (GAI) can efficiently integrate diverse information sources such as images, text, videos, and speech to achieve cross-modal information generation and comprehension. Through long-term training using hundreds of billions of characters extracted from internet data, GAI has approached human intelligence levels. In March 2022, the General Office of the CPC Central Committee and the State Council issued the "Guidelines on Strengthening Scientific and Technological Ethics Governance," designating AI as a key focus area for ethical regulation while proposing clear governance requirements and principles. In July 2023, seven departments including the Cyberspace Administration of China jointly released the "Interim Measures for the Management of Generative Artificial Intelligence Services," charting the course for technological ethics governance. In September 2023, ten departments including the Ministry of Science and Technology jointly issued Article 2 and its appendix of the "Measures for Scientific and Technological Ethics Review (Trial)," establishing a specialized ethical review system for AI activities requiring re-examination under quasi-legislative significance. [9]

In a June 2023 report, McKinsey, a global consulting giant, said that the use of GAI in 63 applications it studied would add $2.6 trillion to $4.4 trillion to the global economy each year, roughly equivalent to the UKs annual GDP. [10]

In March 2024, the United Nations General Assembly adopted its first global resolution on artificial intelligence (AI), calling for the development of "safe, reliable, and trustworthy" AI systems to advance sustainable development. This resolution provides crucial guidelines and governance frameworks for responsible and ethical AI utilization worldwide. The principle of intergenerational justice—ensuring fairness, sustainability, and participation—has emerged as the driving force behind responsible AI practices. As the information age rapidly evolves into intelligent computing, breakthroughs in AI technology continue to emerge, increasingly empowering diverse industries and establishing AI and data elements as prime representatives of new productive forces. [11, 12]

The development of artificial intelligence relies on advancements in algorithmic systems, currently comprising five major schools: symbolic algorithms, connectionist algorithms, Bayesian algorithms, evolutionary algorithms, and analogical algorithms. While these specialized approaches target specific domains and problems, GAI (General Artificial Intelligence) requires algorithms capable of addressing diverse tasks across broad environments. Such algorithms would not only solve the aforementioned specific challenges but also tackle issues in extensive domains – mirroring the ultimate algorithm envisioned by Pedro Domingos. Although this comprehensive solution remains hypothetical, Knud Thomsen has proposed the Ouroboros model as a foundational framework for General Artificial Intelligence (AGI). This innovative cognitive architecture "encompasses all aspects of natural intelligence and AI, including perception, memory, action, learning, reasoning, creativity, and self-awareness." [13].

Chinas artificial intelligence (AI) development has achieved remarkable progress, steadily advancing toward the realm of general AI. While ChatGPT-4 made significant breakthroughs in deep learning, large language models require massive external data resources for training—primarily through external inputs to achieve intelligence. However, this approach has inherent limitations. Human intelligence fundamentally relies on the brains ability to process vast amounts of external information flexibly, primarily due to its complex internal dynamics. This intricate structure enables efficient multitasking. To address this, a research team from the Institute of Automation at the Chinese Academy of Sciences, Tsinghua University, and Peking University has developed a brain-inspired network model centered on "internal complexity." This model aims to simulate neural neuron complexity for more efficient and versatile intelligent processing. The emergence of such brain network models will undoubtedly propel the realization of General Artificial Intelligence (GAI).

The ethical risks of General Artificial Intelligence (GAI) constitute a complex issue, with current academic research primarily focusing on its applications in data, algorithms, and computational power. For instance, GAI poses risks of bias and discrimination at the data level—algorithmic biases stemming from flawed data sources undermine decision-making fairness. Irregularities in data scraping and cross-border transmission may compromise privacy security. At the algorithmic level, GAI presents black boxes and lack of explainability, where opaque systems hinder oversight and evaluation, while algorithmic biases persist due to data quality issues or human annotation flaws. In practical applications, GAI faces risks of misinformation dissemination that disrupts information ecosystems, misleads public perception, challenges human agency, and affects cognitive functions, emotional well-being, and productivity. Current countermeasures include technical improvements (e.g., enhancing algorithm transparency and optimizing data management), ethical frameworks (such as establishing industry ethics standards and promoting value alignment), and regulatory enhancements (including strengthening oversight mechanisms and refining legal frameworks). Beyond theoretical debates, real-world implementations of GAI continue to expose systemic vulnerabilities. [14-19]

In August 2023, Retools client suffered a $15 million loss due to a Deepfake phishing attack. In August 2024, Microsoft was exposed for Copilot containing 15 vulnerabilities, including malicious behaviors such as user impersonation manipulation, authentication bypassing, and data leakage. Despite Microsofts efforts to mitigate risks, no viable solutions have been implemented. Within the ethical framework of GAI technology, its "top-down" generation model lacks social consensus. Self-regulatory and external regulatory mechanisms prove ineffective in algorithmic societies, causing ethical norms to become mere "paper regulations." By contrast, legal enforcement through coercive power and deterrence can clearly define boundaries for technological applications and stakeholder behaviors, adjust the "self-regulating order" of ethics, establish algorithmic preset normative frameworks, and guide the formulation of specific behavioral rules. This enables the legal governance of GAI technologys ethical risks. [20-22].

The current "hard problem" of consciousness primarily concerns the qualitative dilemma of human experience, where subjective perception remains difficult to explain. Given the profound correlation between quantum mechanics and consciousness, some scholars propose that in-depth research into quantum mechanics could elucidate mental processes and establish models to resolve this fundamental question. Brain-computer integration (BCI) represents a promising emerging technology aimed at creating hybrid brain-machine consciousness systems. Numerous research teams worldwide are advancing BCI technologies, exemplified by Tsinghua Universitys Luoping Shis team developing heterogeneous fusion-based brain-inspired computing chips, which are expected to accelerate progress in artificial general intelligence (AGI).

2. The Causes and Realistic Manifestations of Ethical Risks in Generative Artificial Intelligence Technology

2.1. Causes of Ethical Risks in Generative Artificial Intelligence Technology

As a disruptive technology reshaping society, generative AI has emerged as a new engine for economic and social progress, yet it harbors profoundly complex ethical risks. Only through precise analysis of the root causes behind these risks can we develop effective governance strategies. Risk is an inherent characteristic of technology—every technological advancement inherently carries uncertainty and potential hazards.

This dual nature of risk embodies both objective reality and subjective interpretation. [23]

2.1.1. The Development of Machine Consciousness and Quantum Mechanics: Or Leading to Ethical Risks in Science and Technology

Symbolic theorists maintain that computers can establish a "physical symbol system" capable of simulating the brains information processing patterns. While this is merely functional imitation, they claim computers are intelligent and have achieved a simulation of human intelligence. Symbolic representationism, however, reduces human intelligence to mechanical mimicry, oversimplifying it and treating the human mind as computer programs. This approach attempts to grant computers the same intelligence as humans (i.e., strong AI). John Rogers Searle launched a scathing critique of this "strong AI" concept, raising consciousness questions about artificial intelligence: the Strong AI paradox and the Chinese Room paradox. He also demonstrated through the Chinese Room experiment that achieving strong AI via symbolicism is impossible. Although Searles Chinese Room argument primarily aimed to refute the notion of computers possessing minds—and its reasoning itself lacked rigor—it sparked profound contemplation among AI experts and philosophers regarding whether strong AI could possess consciousness.

As a form of advanced artificial intelligence, the ultimate goal of GAI is to achieve human-like intelligence that can adapt to external environments and independently solve various problems including novel challenges. The intelligence demonstrated by humans in problem-solving originates from their conscious mental processes, which means AI development must begin with understanding human consciousness. Therefore, to realize GAI, we must fundamentally question whether machines can develop such consciousness.

Darid John Chalmers posits: "There appear to be no fundamental barriers to the realization of artificial intelligence rationality." In his work *The Conscious Mind*, he describes human brain neural activity as a Combinatorial State Automaton (CSA) – a physical system that replicates the functional organization of neurons. This system contains a functional structure that mirrors the hierarchical organization of brain neurons. Consequently, "according to the principle of organizational invariance, the experiences inherent in this system are indistinguishable from those experienced by the brain itself." [24, 25]

Alongside criticisms of "symbolic computation," scholars have proposed artificial neural networks—models simulating the human brains neural architecture. This connectionist paradigm maintains that the brain processes information through parallel distributed processing, mirroring neuronal connections. While demonstrating learning capabilities and overcoming symbolicisms limitations in information processing, these networks more closely resemble biological cognition. Fundamentally, however, they remain within the framework of computational representationism and "cognitive computability" research. Consequently, even if perfectly replicating the brains neural network, such models would still be mere computational simulations incapable of flexibly addressing novel challenges or achieving universal applicability.

Current research indicates that symbolic computation and artificial neural networks cannot achieve machine consciousness, leaving us to rely on biological neural networks, quantum computing, and brain-computer integration. Both approaches lack biological attributes and are silicon-based materials, whereas the human brain is fundamentally carbon-based and inherently analog. This analog process would symbolize all brain-generated information for computational processing. However, in practice, many aspects of our lives cannot be symbolized or computable. Particularly regarding human consciousness, its intentionality, uncertainty, and emergence cannot be realized through computation. The integration and development of computer science and cognitive science, coupled with deeper understanding of the brains biological structure and functions, have led to the proposal of constructing Biological NeuralNetworks (BNNs) – networks composed of biological neurons, cells, and synapses. By focusing on biological neuron networks and the entire brains neural architecture, BNNs offer a viable pathway for generating biological consciousness.

Specifically, the symbolic paradigm primarily refers to artificial intelligence methods that simulate human brain functions. Symbolists maintain that human cognition and mental processes are computable, treating the brain as an information processing system. They propose that "the brains cognitive processes are essentially symbolic representation and computation." Through philosophical refinement, computationalism has evolved into a crucial theoretical framework for understanding the nature of cognition and the mind. [26, 27]

In recent years, as quantum mechanics has gained traction and evolved, researchers have discovered striking parallels between consciousness generation mechanisms and quantum principles—particularly the uncertainty inherent in consciousness mirrors the uncertainty of quantum motion. Furthermore, concepts like "quantum coherence," "quantum superposition," "quantum entanglement," and "quantum collapse" correspond to descriptions of consciousness dynamics. This profound correlation has led scholars to explore methods for achieving machine consciousness through quantum mechanics. Nobel laureate in Physics Roger Penrose noted: "Consciousness is a phenomenon that cannot be fully explained by classical physics alone. Our mental experiences may stem from peculiar yet fascinating properties of the physical laws governing our world, rather than merely reflecting the structural characteristics of classical physics. In a sense, this explains why, despite the richness and mystery of classical cosmological theories, sentient beings must exist within the quantum realm rather than a purely classical one." [28]

2.1.2. Man-Machine Relationship: Alienation and Inherent Defects of Technology

The persistent concern that "technological progress breeds complacency" has long been a hidden danger. Throughout modern technological history, the social application of any new technology inherently carries highly uncertain ethical risks. These risks stem not only from the objective nature of technology itself but also from human subjective perceptions and societal constructs. Humanity continually adjusts the relationship between society and technology while seeking effective strategies to mitigate these risks, building upon our understanding of technological development patterns. The ethical risks associated with generative artificial intelligence (AI) originate from both the inherent characteristics of this technology and the foundational decisions made in its development.

The instant and comprehensive nature of AI in information provision diminishes peoples ability to filter and judge information, discourages active verification of facts, and weakens critical thinking skills. The convenience and efficiency of AI in information processing lead individuals to rely heavily on it when facing tasks requiring creative thinking, reducing their motivation for independent exploration and innovation. Furthermore, the widespread use of AI in social and emotional interactions causes people to prefer communicating with AI systems over building meaningful connections with real-life individuals, ultimately eroding their capacity to establish deep interpersonal relationships. [29, 30]

In particular, General Artificial Intelligence (GAI) excels at replacing repetitive, tedious, and computation-intensive mental tasks, thereby intensifying reliance on automation and potentially causing skill loss. For low-skilled workers, the urgent need for "re-skilling" to adapt to evolving labor market demands becomes particularly pressing. The capitalist application of GAI undoubtedly exacerbates worker alienation, indirectly increases work intensity, weakens workers control over the labor process, and leads to feelings of their value and capabilities being overlooked or diminished when using these technologies. [31]

Driven by comfort-seeking tendencies or habitual inertia, people increasingly accept machine decisions while relinquishing their own judgment, creating a worrying trend where "humans become mere appendages of machines." As Marx observed: "Humanitys essence is not an abstract quality inherent in isolated individuals. In reality, it constitutes the totality of all social relations." [32]

"Labor produces wisdom, yet it breeds dullness and mental stagnation in workers"; "The more energy workers expend through labor, the more formidable becomes their self-defying world of alienated objects they create". The emergence of GAI has progressively eroded humanitys agency, transforming people into passive participants within capital and mechanical systems. This transformation reshapes human self-awareness and perception of reality, challenges cognitive boundaries, and blurs the traditional distinctions between humans and machines. [33-35]

The rapid iteration of generative AI technology and its "black box" effect introduce uncertainties in technical outcomes and complicate risk assessments. Generative AI models typically feature highly complex architectures with massive parameter scales, making their internal operations and mechanisms difficult to comprehend intuitively. Reverse engineering or explaining their working principles proves particularly challenging, inevitably creating a "black box" effect. This opacity makes it hard to accurately understand how model parameters influence behavior, while also complicating reliability and robustness evaluations. Under such opaque interactions, models become prone to leaking sensitive information or generating false or socially biased content during data processing. Generally speaking, "the controllability of technology is inversely proportional to its risk level – the harder it is to control, the higher the risks." While large language models and thought chains grant generative AI logical reasoning capabilities, they simultaneously make their outputs increasingly unpredictable. [36]

The rapid evolution of generative AI technology has created a dual challenge: while the public remains inadequately informed about existing risks, emerging issues and uncertainties continue to multiply. A prime example is the Sora video model, which has revolutionized short video platforms, advertising industries, and movie trailers. However, its opaque decision-making process raises accountability concerns. When problematic videos emerge, it becomes difficult to identify responsible parties, potentially triggering complex ethical dilemmas. [37]

The relatively lagging ethical regulation of generative AI has weakened the predictability and controllability of technological development. This "double-edged sword" effect means that while humans use technology to enhance efficiency and convenience, they must also continuously explore the laws governing technological development to mitigate risks. Technological innovation requires considering its social impacts through an ethical lens. The deeper humanitys understanding of technological foresight and ethical anticipation becomes, the more effectively we can control technological trends and their consequences. Currently, some countries and regions have established strict privacy regulations, while certain governments and non-profit organizations are developing guidelines and standards to promote fairness and justice in AI systems.

For example, U.S. senators proposed the "2023 Artificial Intelligence Research, Innovation and Responsibility Act", while representatives from the 27 EU member states unanimously adopted the text of the "AI Bill" on February 2, 2024, marking a significant step in the EUs legislative regulation of artificial intelligence. China has successively issued policy documents such as the "Opinions on Strengthening the Governance of Science and Technology Ethics" (2022) and the "Interim Measures for the Management of Generative AI Services" (2023), which have played a crucial role in pro-

moting technology for good. However, as generative AI remains in its early developmental stage with rapid iteration and continuous emergence of new issues, the ethical regulatory system and accountability mechanisms still require further improvement. "As one of the complex technological systems, the algorithms and data behind AI technology are extremely intricate, and humanity has yet to effectively resolve the uncertainty in algorithmic logic." The balance between social benefits and risks brought by technology remains unclear, and ethical values and legal adaptability corresponding to technological development need to be established and refined. Humanity still lacks a clear and comprehensive understanding of the consequences of generative AIs societal applications. Moreover, the application and proliferation of generative AI are global in nature, requiring coordinated efforts from all countries in ethical risk governance. However, "different nations have varying strategic interpretations of artificial intelligence."The disagreement on the goal is creating the division of the international community and undermining international cooperation, "which will reduce humans ability to control and respond to the ethical risks of generative AI and the uncertainty of the consequences of ethical risks. [38-40]

The subjective perception of ethical risks in generative AI strengthens judgments about their likelihood. Generally, risks arise from both objective causes and subjective factors. The subjective understanding of technological risks is influenced by social environments and mass media, while the cognitive mechanisms and strategies of risk-perceiving individuals may also affect technical entities perception of such risks. Regarding ethical risks in generative AI, peoples level of familiarity with the technology may influence their perception of its ethical hazards. Technicians with better understanding of the technologys operational mechanisms and development trends might be more aware of potential risks like privacy breaches, data bias, and misleading information generation, whereas those less informed may lack sufficient awareness. Differences in cultural backgrounds, religious beliefs, personal circumstances, and educational experiences can lead to variations in how people perceive ethical risks. Factors such as emphasis on privacy rights, understanding of fairness and equality, and respect for individual autonomy all impact subjective perceptions of generative AIs ethical risks. Moreover, stakeholder interests and risk perception biases may further influence ethical risk assessments. Public or government agencies, driven by concerns for public welfare, might demonstrate heightened sensitivity to ethical risks. While recognizing the importance of ethical considerations, generative AI innovators may underestimate ethical risks due to their technological advantages. Tech companies might even combine capital resources with algorithms to maximize profits, extensively collecting user data and digital footprints to build models that enable behaviors like price discrimination and labor exploitation—actions that trigger ethical hazards. Ultimately, peoples subjective perception of risks forms the structural basis for ethical risk formation, significantly influencing risk assessment and governance decisions. [41, 42]

2.1.3. Information and Content Sources: Limitations and Unconscious Mirroring

When GAI processes and generates information, its ability is limited by the size of the model training data set and the time period. At the same time, the "black box" nature of the algorithm means that people cannot ask and solve the logic behind the decision. [43]

The countermeasures proposed by GAI may be uncritically accepted by users, posing significant ethical risks. Firstly, the effectiveness of GAI largely depends on the quality and scope of its training data. The lack of diverse datasets results in insufficient generalization capabilities for models under specific scenarios, compromising their practical applicability. Secondly, while GAI can generate smooth and seemingly reasonable text, it cannot truly comprehend the actual meaning behind the generated content, making it prone to unintentionally producing inappropriate or offensive material. Thirdly, when generating lengthy structured content, GAI may repeat previous viewpoints, failing to effectively organize and structure information. As Freud noted, the unconscious aspect of human psychological activities encompasses instinctual impulses, desires, and repressed emotions. Analyzing these unconscious contents can reveal the true motivations behind individual behaviors, aiding self-awareness and psychological balance. [44]

In contrast, the unconscious concepts within GAI can be likened to the "id," "ego," and "superego" in intelligent systems. The "id" represents GAIs fundamental programming code and instinctive impulses—raw data and algorithms processed without filtering. The "ego" serves as GAIs operational interface, handling logical decision-making while balancing internal needs with external environments. The "superego" governs ethical algorithms and social programming, guiding GAIs moral responses. Operating through preset algorithms and programs, GAI lacks subjective experience or self-awareness, unintentionally replicating patterns and structures from its training data—especially when no explicit instructions prevent such behavior. The essence of GAIs operation lies in analyzing and learning vast amounts of data. Its designers and users aim to avoid unintended mirror-like reactions. [45]

Error Sources and Value Biases in GAI: GAI systems are trained on large open-source datasets ranging from millions to trillions of data points. While these datasets help models understand language patterns and answer questions, they often contain biased content. As GAI learns from such data, it may internalize these biases and replicate problematic elements in subsequent generation tasks. Since GAI is trained on specific datasets, theres no straightforward way to trace outputs back to their original inputs or ensure the model "overwrites" potentially problematic, sensitive, or illegal data. This creates vulnerabilities for hallucinations, misleading content, and factual errors in generated outputs. Although large lan-

guage models like GPT-4 and DeepSeek have improved accuracy in producing authentic answers, they still face the "AI hallucination" issue—mistaking false or misleading information for factual truth. Functioning similarly to advanced auto-completion tools, GAI predicts next-word sequences based on observed patterns. Its primary goal is to generate credible content rather than verify its authenticity. [46]

Even when GAI models are trained using accurate data, their generative nature means they can still produce new, inaccurate content through unintended means. As the saying goes, "Theres no such thing as a panopticon perspective." When evaluating any subject, GAI imposes a specific reference framework, making it inherently incapable of neutrality or objectivity. The operational mechanisms of GAI may dismantle the established perceptions and values among social members, potentially leading to weakened ideological leadership. [47]

2.1.4. Limitations of Technology Itself: Algorithms and Computing Power Create New Forms of Digital Inequality

"Technology is out of control not because technology itself is autonomous and evil, but because we ignore our understanding and responsibility for technology." [48]

The ethical risks of generative AI stem from the imbalance between instrumental rationality and value rationality. From the perspective of technological instrumentalism, the blind pursuit of efficiency and effectiveness has created a conflict between digital judicial technology and judicial values, manifesting as an overemphasis on "technology-centric" approaches while neglecting or diminishing "human-centered" principles. This imbalance primarily manifests through excessive reliance on artificial intelligence technology and blind worship of its potential, as well as the erosion of the subjective significance of judicial professionals. The promotion of instrumental rationalism emphasizes AI technology as a means to maximize efficiency and gain benefits, thereby overlooking value rationalitys emphasis on human dignity and fundamental needs. This not only leads to technological alienation but also triggers ideological risks: the capitalistic and ideological attributes inherent in AI technology unconsciously shape and guide peoples ideologies and values. On the other hand, overemphasizing technologys instrumental nature fosters unhealthy competition and conflicts, neglecting the importance of benign technological development and proper application. This results in contradictions between humans and technology, as well as among people themselves, highlighting a control-and-be controlled dynamic that exacerbates interpersonal tensions and human-machine conflicts.

Technology itself has increased the requirements for digital literacy: The rapid development of GAI is reshaping the social information ecology and interaction mode, and puts forward higher multi-angle requirements for digital literacy.

First, GAI provides complex information, requiring people to improve and strengthen their ability to understand information, so as to use critical thinking to distinguish between true and false; screen out reliable and key information, that is, the dimension of information processing improves the requirements of digital literacy.

Second, the new core elements of digital literacy are that there are many privacy, security and ethical risks in the process of data processing, so it is necessary to enhance the awareness of data privacy and ethical security.

Third, the application of technology needs to raise the expectation of digital literacy, because the application of many functions of GAI forces people to master the operation skills and have the corresponding innovative application ability.

Fourth, humans need good digital collaboration and communication skills because of the increasing number of scenarios in which they interact with AI systems.

Fifth, the continuous and rapid evolution of artificial intelligence technology inevitably requires people to engage in uninterrupted learning and sustained adaptation. Particularly in specialized fields and scenarios such as healthcare and legal practice, it is crucial for humans to enhance their practical capabilities and develop the competence to deeply understand potential social and ethical implications. The accountability and ethical considerations inherent in General Artificial Intelligence (GAI) demand that humans demonstrate strong social responsibility and moral awareness. It is essential to fully recognize the severity of these issues and strictly adhere to relevant laws, regulations, and ethical standards in practice. This approach helps mitigate ethical concerns related to personal privacy, data security, and intellectual property rights associated with GAI applications.

Humanity must remain vigilant against the massive information monitoring and technological abuse driven by General Artificial Intelligence (GAI). Real-time analysis of text data reveals that GAI can autonomously monitor human online activities, opinions, and interactions without individuals knowledge or explicit consent. For instance, OpenAIs privacy policy states that it automatically uses user-provided data to train ChatGPT models by default. The indiscriminate collection and analysis of personal information clearly undermine individual autonomy over their privacy. The expanded surveillance capabilities enabled by GAI easily normalize societal monitoring, intensifying the constant awareness of being watched—even digital surveillance automatically induces an unconscious self-examination, mirroring Jeremy Benthams concept of the "Panopticon." Meanwhile, GAI remains vulnerable to malicious attacks such as prompt injection and jailbreaking, which can be exploited to obtain sensitive data, spread misleading information, or execute malicious commands.

With the rapid evolution and iterative optimization of cutting-edge technologies in GAIs upstream and downstream sectors, the misuse of these technologies has become increasingly prevalent. Firstly, GAI can generate exceptionally realistic artificial content such as videos and audiovisual materials. When used for malicious purposes like defamation,

fraud, or political manipulation, it poses significant threats to citizens right to reputation and public order. Secondly, user personal information is often misused beyond its intended applications—such as commercial analysis, advertising targeting, or market research. Unauthorized use of personal data violates the legal principle of minimal necessity, eroding users control over their information. Thirdly, GAI aggregates fragmented online personal data to create detailed digital profiles, including sensitive details like preferences, habits, and social connections. These profiles are then exploited for identity theft and illegal commercial activities.

Bernard Stigler, a French scholar, argues that humans are born with "inherent flaws". [49]

To compensate for this human "defect", technology has been developed to create external prosthetics that enable complete human existence. However, these artificial limbs inherently carry primitive flaws since their creation. The inherent limitations of digital judicial technology itself form the internal logic behind its ethical risks and other systemic issues.

The performance of digital judicial systems primarily depends on the quality of model design, training, and debugging. During the design phase, factors like data quality, parameter tuning, and model evaluation can introduce cognitive biases. In the training and debugging stages, intelligent technologies such as image, text, and voice generation rely on integrated techniques including variational autoencoders (VAE), generative adversarial networks (GAN), and recurrent neural networks (RNN). For instance, variational encoders typically utilize two neural networks—the encoder and decoder—to collect and learn latent features from data, analyze and integrate data structures to create deep learning models for generating new samples. As parameters increase or systems undergo iterative updates, more unique yet stylistically similar new samples emerge. Consequently, data quality and quantity determine the performance and practical application of large models. However, data contamination or tampering during training refers to intentionally adding erroneous samples to standard training datasets, which distorts or compromises data integrity, leading to biased outputs that affect final decision-making. Additionally, GAIs massive computational power involves enormous parameter scales. For example, the globally popular GPT-3 was developed using approximately 1750 parameters. GPT-4 uses 1.8 trillion parameters, while the number of parameters is up to 100 billion. Such a huge amount of parameters are mainly obtained by crawler technology to read web information, network dialogue and online book resources, etc. The legitimacy and transparency of crawler technology are directly related to users privacy and data security.

The opaque nature of algorithmic design and formalized logic pose credibility risks in digital judicial systems. Artificial intelligence amplifies the uncertainty inherent in technological development, primarily stemming from the lack of transparency and interpretability in algorithmic decision-making. This cognitive challenge undermines the predictability and assessability of AIs decision outcomes, while practically complicating the governance of ethical issues arising from such systems.

The limitations of resources and technology have led to uneven distribution and the objective existence of technical barriers: Due to disparities in digital resources, educational levels, and economic capabilities, there are significant differences in accessing and applying GAI across different countries, regions, industries, and groups. First, infrastructure disparities. The application of GAI relies on powerful computing capabilities and high-speed internet connectivity. Developed countries and urban areas possess more sophisticated information infrastructure, while developing nations and regions face objective constraints in accessing and using GAI due to inadequate infrastructure. Second, economic capability gaps. The development and deployment of GAI require substantial financial investment, so the economic capacity of enterprises and individual users directly impacts their ability to access and utilize GAI. Third, technological iteration differences. The advancement of GAI necessitates continuous technological updates and maintenance. Users with better economic conditions can keep pace with technological iterations, while those in poorer economic conditions are forced to use outdated technologies due to cost issues. Currently, data, algorithms, and computing power have gradually become new production factors. GAI has first been popularized in technologically advanced countries and regions, with a few economies exerting dominant control. Developed nations and tech giants and enterprises rely on dataBy leveraging their advantages in accumulated resources, computing capabilities, and R&D expertise, developing countries can more effectively utilize Global Artificial Intelligence (GAI) to drive economic growth and significant social progress. However, constrained by resource limitations and other factors, small and medium-sized enterprises (SMEs) struggle to achieve comparable success in GAI-related fields. This disparity will consequently widen the global economic divide significantly.

2.1.5. Regulatory Failure: Lack of Stakeholder Responsibility Awareness

From the perspective of regulatory framework development, existing mechanisms lack sufficient professionalism. The GAI (Generative Artificial Intelligence) corpus construction and update processes suffer from deficiencies in privacy policy disclosure, making it difficult for users to fully comprehend information processing procedures. Meanwhile, data sources are complex and diverse, with some web-scraped data lacking traceable legitimacy and authenticity. Although current regulations address related issues, they fail to clearly define standards for verifying generated content authenticity, leaving ambiguous criteria for determining service providers fault and lacking effective governance frameworks. Furthermore, the complexity of responsible entities creates regulatory challenges, with frequent occurrences of overlapping

oversight responsibilities and bureaucratic buck-passing, compounded by overlapping jurisdictions among multiple regulatory bodies. For GAI systems characterized by massive scale, trillions of parameters, rapid scalability, and extensive application scenarios, traditional regulatory approaches marked by delayed implementation and fragmented enforcement face challenges including sluggish response, insufficient flexibility, unclear regulatory benchmarks, and high cross-domain coordination costs. [50]

From the perspective of regulatory applicability, informed consent faces challenges in implementation. On one hand, individual control rights are effectively undermined as users lack clarity regarding the scope and extent of GAIs data collection and usage. The ambiguous boundaries of data utilization make it difficult to strictly enforce the purpose limitation principle, often leading to unauthorized data usage that compromises user rights and interests. [51]

On the other hand, inadequate disclosure of privacy policies has led to most GAI platforms failing to adequately inform users about critical information regarding the scope, methods, and purposes of data processing. This results in superficial user consent that becomes merely formalistic. From a governance perspective, prominent issues include insufficient judicial remedies and regulatory gaps. The inherent concealment, latency, and complexity of GAI infringement make it challenging to establish and quantify standards for personal information damage, thereby continuously undermining the relevance and effectiveness of judicial remedies.

Due to the profit-driven nature of Generative Adversarial Intelligence (GAI) in technological development and the lack of regulatory oversight, attackers can compromise GAIs reasoning and predictive capabilities by injecting malicious samples into training datasets or altering label information. Malicious actors exploit the systems reliance on data by inputting manipulative commands and erroneous information, guiding GAI to learn inaccurate or harmful behavioral patterns that potentially influence its decision-making. Errors and inaccuracies accumulated during interactions are unintentionally learned and amplified by GAI, spreading misinformation or harmful behavioral patterns across broader social interactions.

Under the influence of human errors, Generative Artificial Intelligence (GAI) generates false misinformation that distorts public perception of content. Three primary mechanisms contribute to this issue: First, algorithmic design flaws may introduce specific biases into the system through intentional coding. Second, humans naturally search for, interpret, and retain information with confirmation bias, creating sample processing bias during GAIs development and use. Third, developers unintentionally impose their perspectives on models or train GAI using biased datasets, both contributing to cognitive biases in AI systems.

In social systems, moral responsibility does not exist in isolation but emerges as a product of relationships, embodying the ethical expectations society imposes on moral actors. Responsibility constitutes "a fundamental normative relationship whose foundation may reside not only in morality and law, but also in conventional or role-specific regulatory requirements." [52]

Artificial intelligence (AI) is a complex artificial system whose technical principles and social consequences determine the allocation of governance responsibilities and the selection of governance approaches, involving multiple stakeholders. In the application and governance of AI in digital judicial systems, the lack of awareness regarding different principal responsibilities and the absence of responsible conduct are key factors contributing to ethical risks.

There exists a risk of generating and spreading biased and discriminatory content. With the innovative application of digital and information technologies, social biases and discrimination can now emerge and spread more conveniently and rapidly through technological means compared to traditional eras, with their negative impacts becoming increasingly widespread. The emergence of computer algorithms has particularly turned algorithmic discrimination into a prevalent ethical concern in technology. While generative AI empowers peoples production and daily life while improving work efficiency, it may significantly amplify the risks of generating and spreading biased content, often in more concealed forms. The natural language models underlying generative AI are built on algorithms and massive datasets. If developers incorporate personal preferences or specific biased values into training data, these models may learn and reflect such biases, leading to socially unequal, prejudiced, or discriminatory content and decisions. In other words, "if the training data used to develop AI algorithms and models contains biases, the algorithms themselves will also be biased, resulting in discriminatory outcomes in the responses and recommendations generated by generative AI." [53]

2.1.6. Institutional Failure: The Construction of Institutional Norms for the Application of Artificial Intelligence Lags Behind

The refinement of policy frameworks demonstrates a forward-looking governance approach to AI technology applications, where risk prevention takes precedence over problem-solving. This reflects the characteristics of holistic planning and systematic design. Effective institutional development can address the "lack of teeth" in educational AI ethics principles—where ethical guidelines may establish normative ideals but lack mechanisms to enforce compliance, and violations go unpunished. [54]

To safeguard the public relations objectives that underpin ethical principles from being eroded by economic interests, comprehensive institutional frameworks for artificial intelligence must be implemented throughout its research and application processes. The development of inadequate regulatory systems could ultimately hinder AIs technological progress.

It should not be ignored that the development of technological practices is mostly preceded by the formulation of policies and regulations. With the rapid change and update of artificial intelligence technology, there are still many institutional loopholes in its technical supervision and application norms.

1) First, China has yet to establish comprehensive legal frameworks specifically targeting artificial intelligence in the digital justice sector. To regulate the effective and rational application of AI technologies, the country has implemented a series of guidelines including construction standards, action plans, and policy recommendations, supported by relevant government incentives that chart the course for AI development. While Chinas judicial system has introduced regulations to standardize blockchain applications and AI in legal contexts, there remains a critical gap: the absence of systematic legal provisions capable of addressing the complex and ever-evolving practical scenarios encountered in real-world operations.
2) China has yet to establish unified ethical guidelines for AI-powered judicial applications. The European Unions "Ethical Guidelines on Trustworthy Artificial Intelligence" outlines four core principles: "well-being," "non-maleficence," "autonomy," and "fairness with explainability." China has introduced its "Ethical Standards for Next-Generation Artificial Intelligence," emphasizing regulatory frameworks, supply mechanisms, R&D protocols, and usage guidelines. However, existing principles and regulations still face challenges such as ambiguous definitions and limited practical effectiveness in implementation.
3) In the process of artificial intelligence and digital judicial application in China, there is no formal special regulatory agency, so there is a lack of targeted and scientific regulatory system, so it is objectively difficult to implement supervision and management of different subjects behavior.

2.2. Ethical Risk Characterization of Generative Artificial Intelligence

In recent years, both domestic and international players have invested substantial resources in artificial intelligence to dominate the information technology landscape. While GAI (Generative Artificial Intelligence) was initially a theoretical concept, the emergence of ChatGPT-4 has demonstrated remarkable natural language processing capabilities and learning abilities. This breakthrough has paved the way for implementing "big data models + multi-scenario" architectures within GAI frameworks. Some scholars even argue that certain generative AI systems have already reached the preliminary stage of general artificial intelligence. [55]

The emergence of Sora, in particular, has achieved a qualitative leap in natural language understanding and processing capabilities, prompting many scholars to remark that the era of "General Artificial Intelligence" is approaching. As Herbert Alexander Simon, the "father of artificial intelligence," once observed: "We now have machines capable of thinking, learning, and creating. Their abilities are growing at an unprecedented rate, and within the foreseeable future, their problem-solving capabilities will keep pace with human cognition." [54]

Simons optimistic prediction now appears to be gradually materializing. As the world transitions into an era of digitalization and information technology, scientific and technological advancements are profoundly shaping societal development. The evolution of artificial intelligence (AI) remains unstoppable—both its technological progress and the exploration of machine consciousness are advancing toward General Artificial Intelligence (GAI) and even Artificial General Intelligence (AGI). However, AI systems that rival human intelligence in cognition, consciousness, and emotional capacities will inevitably trigger ethical risks for GAI, including enhanced moral agency, absence of accountability, and emotional subjectivity misalignment.

Artificial intelligence has driven profound transformations in digital justice through its disruptive innovation capabilities, driving effective improvements across judicial entities, subjects, environments, and intermediaries. However, as an advancing technological force, humanity still struggles to fully grasp the practical applications of AI. With the continuous upgrading and deepening implementation of digital judicial technologies, ethical risks arising from cognitive dimensions, subject-related aspects, knowledge domains, and data dimensions are escalating. Only through rational understanding and clear characterization of these ethical risk patterns can we establish a solid foundation for better regulating AI-driven digital justice systems.

While generative AI is revolutionizing our technological landscape with its unique creativity, interactivity, and versatility – profoundly reshaping humanitys relationship with reality – it remains in its early developmental phase. The systems evolving capabilities create complex social impacts, carrying significant uncertainty and posing substantial ethical risks that demand careful consideration.

The term "risk" generally refers to potential harmful events and their probabilities. Risk is closely related to uncertainty, typically encompassing both the unpredictability of outcomes and the volatility of occurrence probabilities. Ethical risk specifically concerns the likelihood of negative consequences arising from ethical and moral dimensions. The uncertain ethical repercussions generated by AI technology development and application constitute ethical risks, manifesting as imbalanced ethical relationships, social disorder, and behavioral misconduct. [56, 57]

The ethical risks of generative artificial intelligence (AI) refer to the potential uncertainties and possibilities that may lead to negative ethical consequences during its development and application. These represent unintended outcomes that

humanity does not anticipate when innovating with GAI. The primary ethical risks associated with generative AI can be summarized as follows:

2.2.1. Cognitive and Subject Dimension: The Intervention of Digital Virtual Body Lead to the Decline of Judicial Workers Subjectivity

When the application of GAI demonstrates a trend of surpassing in numerous fields, the technological approach may deviate from its original correct trajectory and even escape the control of the subject to some extent, thereby forming an ethical colonization of subjectivity. When information technologies centered on artificial intelligence are externally implanted into judicial practices, technocentrism and algorithmic justice objectively render mathematical algorithm libraries incapable of adapting to the complex and ever-changing scenarios in Chinas judicial practice activities.

Artificial intelligence follows the formalized and standardized operation logic. On the "Procrustes iron bed" of technology, the workers of the judicial professional community are uniformly symbolized and materialized, reduced to an object that can be calculated, and their active thinking and spiritual freedom will be weakened or even lost.

While AI provides judicial practitioners with precise recommendations, it also fosters psychological dependence on technology. Blindly treating algorithmically generated information as "divine revelations" risks dulling judicial professionals critical thinking and moral compass. A Boston Consulting Group study on how generative AI creates value and destroys it reveals striking patterns: participants using ChatGPT-4 for idea generation produced 41% fewer diverse viewpoints than non-users. Moreover, after receiving suggestions from ChatGPT-4, respondents showed little inclination to expand their perspectives with more varied viewpoints. [58]

For instance, embedding hidden codes containing racial or gender biases in algorithms may introduce covert discriminatory tendencies into generated content. This could manifest as social stereotypes and prejudices against specific groups within the text, images, or other outputs. Such biases often remain undetected in training datasets, making it challenging for models to identify and correct these underlying issues. Consequently, generative AI systems can easily spread and amplify such prejudices through widespread usage. A typical example is when users request images of high-paying white-collar jobs – the AI-generated portraits almost invariably depict predominantly Caucasian individuals. [59]

This phenomenon can manifest across various content-generation domains including text, images, audio, and video, potentially shaping user perspectives through biased narratives. The concept of intersubjectivity – also termed interactive subjectivity or mutual subjectivity – emphasizes "the interconnectedness between individuals as subjects and others in objectifying activities." [60]

German philosopher Jürgen Habermas elucidated intersubjectivity through the lens of social communicative action, conceptualizing it as a process enabling individuals to establish and develop meaningful mental exchanges. This theory not only affirms human agency but also fosters interactive engagement between people, thereby transcending the self-centered individualism inherent in subject-oriented philosophy. It advocates for equitable, harmonious interactions characterized by mutual respect and understanding. As Habermas noted: "In the educational process, educators function as subjects while learners are treated as objects; conversely, in the learning process, learners become active agents whereas educators serve as passive recipients." [61]

In modern education, the emphasis on intersubjectivity has replaced traditional one-way "cramming" teaching methods. Educators and learners now interact through mutual respect built on equality, democracy, communication, and understanding, with a growing recognition of each others intellectual worth. This shift moves education from the conventional "teacher-student" model to a collaborative "teacher-student-teacher" framework that respects students developmental patterns. However, as digital lifestyles dominate modern existence, artificial intelligence technologies—once tools for mutual understanding and co-construction of knowledge—have increasingly become the dominant force shaping teacher-student interactions. This technological dominance has gradually eroded the essential intersubjectivity between educators and learners.

Regarding the ethical status of artificial intelligence, there are currently three main views in Chinas academic circles: First, artificial intelligence is only an ethical object; second, artificial intelligence is an artificial subject; third, artificial intelligence is a super artificial subject.

According to traditional anthropocentric views, artificial intelligence cannot become a moral agent regardless of its development. Hubert Lawrence Dreyfus and Susan Searle argue that general artificial intelligence can only exist as a moral object rather than an agent, since only humans with embodied perception and cognition possess the capacity to judge and perform moral actions. As AI is fundamentally a computer program, it cannot match the human mind. Thus, while such advanced AI might exhibit similar moral behaviors, it lacks the capacity to comprehend them, remaining strictly an object of morality. From an anthropocentric perspective, this conclusion holds unshakable validity: by definition, artificial intelligence can never attain the biological foundation of human intelligence—the "mind." International scholars Luciano Floridi and J. W. Sanders contend that AI represents "mindless morality," where agency is primarily manifested through the interactivity, autonomy, and adaptability of intelligent systems, without requiring "free will or mind." [62]

"Deed facts and ethical meaning exist in the network formed by the interaction of actors, and the definition of actors ethical identity must also be carried out in this network of mutual relations." [63]

Therefore, in practical applications, artificial intelligence

can be termed an "artificial subject" whenever it exhibits interactive and autonomous response behaviors. Functionally, these entities primarily perform specific tasks and operations such as autonomous driving and smart manufacturing. Since they operate based on predefined algorithms and rules, they lack self-awareness and moral judgment capabilities.

Generative artificial intelligence (AI), as a representative of general AI, has made continuous advancements in deep learning and natural language processing. These developments have yielded various achievements that are being applied and influencing human society, prompting scholars to increasingly focus on the ethical implications of general AI. Joseph E. Nadeau, an American information technology expert, posits that general AI will evolve into "super artificial entities" in the future. [64]

The reason for his prediction lies in his belief that generative AI makes all actions and judgments based on specific logic, algorithms, and procedures, thus constituting an absolutely rational entity. In Kants view, moral subjects are rational beings who follow ethical laws. Regarding rational capabilities, generative AI clearly better meets the requirements of moral subjects. Current scientific discussions about super artificial entities remain at the realm of science fiction. Through exploring the definition and potential of generative AI, it becomes evident that such systems are "capable of executing universal functions including image/voice recognition, audio/video generation, pattern detection, question answering, translation, while also possessing multiple intended and unintended purposes." [65]

2.2.2. Knowledge and Data Dimensions: Academic Integrity, Knowledge Oligarchy and Privacy Leakage

The birth of ChatGPT has triggered "a revolution in super-language, cross-media and multimodal content generation and intelligent technology". [66]

As a conversational large language model, ChatGPT can generate structured content based on users value expectations, enabling natural language responses to user inquiries and even providing satisfactory answers tailored to specific needs. While its demonstrated strong interactive capabilities and generative power continue to drive continuous innovation in scientific research paradigms, issues such as academic misconduct and knowledge monopolization have gradually emerged.

Academic integrity risks and ethical misconduct in research: The widespread use of generative AI may lead to ethical violations, complicating the governance of research misconduct. Since generative AI produces content that appears plausible but is actually fabricated, it could be exploited to create false research outcomes or data. This results in inaccurate conclusions and the spread of misinformation, undermining the objectivity and credibility of scientific research. Moreover, when models generate content involving personal privacy information without proper informed consent, it violates academic principles of informed consent, invalidating research activities. If researchers use generative AIs data transcoder feature to convert original texts from others works into altered structures and language, they can create new content as their own work. "Not only can people leverage generative AI for data collection across vast databases, but they can also continuously refine text expressions and content through interactive feedback. Users might even instruct generative AI to write papers in specific styles, making the articles mimic their own writing style with such precision that they appear indistinguishable from authentic works." [67]

When these "deep fakes" are published publicly or released on the Internet, they will not only bring legal disputes over copyright ownership, but also inevitably bring great negative effects on social integrity. They may also cause researchers to rely on tools, resulting in shallow thinking ability, cognitive deterioration, and colonial consciousness of capital. [68]

As a form of knowledge representation, artificial intelligence inherently carries ideological attributes. This means that the technological development, rule-making, and widespread adoption of intelligent education systems are primarily controlled by developed countries or regions. However, factors such as technological backwardness and legal policy restrictions have left less-developed nations and regions trapped in a state of "information poverty" and "technological underdevelopment." The concept of "mass intelligence connectivity" represents an open vision for integrating knowledge from AI models with human expertise, yet it risks creating its opposite—— a new "knowledge oligopoly" centered on silicon-based knowledge rights——. [69]

Western-originated intelligent models – including ChatGPT, Google, Wikipedia, and major academic databases like Springer – have dominated knowledge acquisition. This dominance has created intellectual dependency in less developed regions, essentially mirroring the cultural exports and capital expansion of Western tech giants. The systemic inequality behind these platforms threatens to undermine global knowledge production systems. [70]

In the data acquisition process, digital judicial users and developers maintain an imbalanced power dynamic. Digital judicial software requires users to upload personal information and may even forcibly access private data, while developers dominate user data collection. This creates a "panoramic surveillance" scenario where all data, information, behaviors, and statuses generated in digital judicial activities are visualized as images and data. With inconsistent data collection standards and inadequate server security measures, unauthorized intruders gain opportunities for data theft, increasing risks of user data breaches. Such vulnerabilities jeopardize the personal safety of digital judicial practitioners, hinder corporate development, and expose individuals privacy. In severe cases, sensitive information leaks may even trigger extreme incidents. [71]

On the other hand, the erosion of privacy boundaries and ambiguous data usage rights in digital justice systems not only

lead to data breaches but also severely violate users "right to data erasure" —the fundamental right of data owners to demand platforms delete unauthorized collected information. However, during the development and application of AI foundational models, when user data (including linguistic culture, religious beliefs, etc.) becomes training "feed," the resulting models inherently retain deep learning outcomes based on such user data. ChatGPTs account cancellation guidelines explicitly state: While users can delete their account information, sensitive data from conversations cannot be removed and will be retained for future training purposes. [72]

Therefore, clarifying the direction and boundary of data use in digital justice is the key to safeguarding the legal rights of data subjects and resolving the risks of data ethics and security.

It brings threats to peoples life, property safety and even personal safety. Moreover, "the data obtained by illegal crawling often has the characteristics of confidentiality, high density and protection, so the illegal behavior not only infringes on individual information rights and interests, but also involves national data security and data sovereignty". [73]

There exists a risk of spreading misinformation and cognitive manipulation. Generative AI, powered by massive computational power to learn from vast datasets, can provide users with suggestions or answers. While people generally trust generative AIs data-driven, seemingly objective probabilistic inferences, the content generated isnt logically reasoned but rather filtered through empirical learning. These models lack the ability to verify the accuracy or authenticity of their outputs, leading to widespread risks of producing false or erroneous information. For instance, errors in training data may cause models to generate corresponding inaccuracies. They might produce answers that appear reasonable yet are incorrect, nonsensical, or fabricated—phenomena known as AI hallucinations. This refers to AI generating content that sounds plausible but is actually false or fictional, creating risks of misinformation dissemination. Such issues could potentially mislead society, resulting in misguided judgments and decisions. [74]

The video generation model Sora could potentially fuel deepfake technology, even posing risks to human safety and property security. NewsGurd, a U.S.-based news credibility assessment and research institution, found through testing ChatGPT that it can alter information within seconds and generate large amounts of convincing yet source-unverified content. When generative AI is deliberately exploited to produce erroneous content, the risk of spreading misinformation becomes increasingly complex. Model developers often intentionally incorporate specific information into training datasets based on particular beliefs, preferences, or purposes. Consequently, GAI inadvertently reinforces certain viewpoints and values in generated content. By producing attention-grabbing posts, comments, and feedback, GAI influences public perceptions on social media platforms, manipulating societal cognition and values. This ultimately undermines cultural diversity and erodes public trust in information credibility, leading to a crisis of confidence.

2.2.3. The Crisis of Human Autonomy: Over-Dependence on Technology to Dissolve the Subject Consciousness of Human Beings

Human subject consciousness refers to the ability of human beings to exist as autonomous and conscious beings, to self-determination, self-reflection and self-creation. This is not only the fundamental feature that distinguishes human beings from other creatures, but also the core driving force for the progress of human society and the development of civilization. [75]

Human self-awareness and identity are fundamentally shaped by personal experiences, cognitive frameworks, and value systems. Over-reliance on General Artificial Intelligence (GAI) diminishes individual growth through limited knowledge accumulation, leading to homogenized values and beliefs that ultimately erode self-efficacy while stifling creativity and personality development. The rapid advancement of GAI has inflated human rationality, revealing modern societys existential paradox: the deeper humanitys dependence on AI, the more susceptible we become to digital technologys control mechanisms, perpetuating a binary opposition between humans and machines.

The continuous advancement of digital technology has increasingly challenged human subjectivity in moral and ethical dimensions. Consequently, human agency will no longer be the sole decision-maker but must negotiate and coexist with digital technologies and artificial intelligence. The entire process of GAI algorithmic power exercise remains opaque, transcending, deviating from, and operating independently of human rights. This results in a gradual erosion of human rights boundaries, leading to the phenomenon of subject-object alienation. The black-box nature of GAI and its reliance on simulated environments directly or indirectly cause humans to blindly accept and trust GAIs outputs. To some extent, this trust inevitably surpasses peoples direct experiences and perceptions in the real world.

Humans are both moral and responsible entities: As moral subjects, they must bear accountability for the consequences of their actions. According to traditional perspectives, humans should be held accountable for the behaviors of artificial intelligence (AI). While AI, like large machines, is a human creation, it possesses inherent intelligent algorithms and learning capabilities. This means humans cannot fully control AIs outputs or clearly describe its entire operational process. Essentially, AI inherently possesses autonomy when handling tasks. This autonomy makes it impossible for designers, manufacturers, and related responsible parties to fully predict the future behaviors of GAI (General Artificial Intelligence) or hold them accountable.

In the early 21st century, Andreas Matthias observed: "Traditionally, manufacturers/operators of machines bore

(moral and legal) responsibility for their operational consequences. Autonomous learning machines based on neural networks, genetic algorithms, and agent architectures have created a new paradigm where manufacturers/operators can no longer predict future machine behavior in principle, thereby losing moral accountability. Society must decide whether to abandon such machines (which isnt a practical option) or face a responsibility gap that traditional liability frameworks cannot bridge." [76]

This has led to growing concerns about artificial intelligences potential ethical responsibilities. However, current AI systems lack the practical capability to assume such obligations, creating a risk of liability absence at this developmental stage. Take autonomous driving technology as an example: "The emerging field of self-driving vehicles recently encountered a real-world trolley dilemma – a classic ethical quandary where an autonomous vehicle approaching another vehicle or pedestrians lacks sufficient time to stop." [77]

Regardless of the choice, accidents remain a potential risk. Autonomous driving systems, being artificial intelligence with inherent autonomy, cannot be fully accountable to their manufacturers and designers. This is because "the modular design of computer systems may mean that no single individual or team can fully comprehend how the system interacts with complex new input streams or respond to them." [78]

In conclusion, since neither manufacturers nor designers can fully comprehend the operational mechanisms and underlying causes of autonomous driving systems in real-world scenarios, they cannot be held entirely accountable for their actions or consequences. While AI-driven learning algorithms fundamentally challenge traditional liability frameworks, it remains unrealistic to hold them responsible for autonomous vehicle technologies. As General Artificial Intelligence (GAI) algorithms exist solely as computational programs without physical embodiments, determining liability subjects proves inherently difficult, making it impossible to assign accident liability to these systems.

The risk of accountability gaps arising from the relatively limited autonomy in developing specialized AI becomes increasingly prominent as we progress toward general artificial intelligence (AI). As AI gains greater autonomy and self-awareness, it can now "flexibly" process and respond to complex scenarios without human intervention. This remarkable ability to handle integrated tasks autonomously makes AIs behavior increasingly resemble human actions, relegating humans to a secondary role. General AI evolves into independent decision-makers and responsible entities, with their actions often determined by real-time environmental interactions that surpass the precise predictability and control capabilities of researchers and manufacturers. [79]

By analyzing human users behaviors and preferences, GAI provides algorithmic recommendations for personalized content. This targeted information delivery leads to homogenization of content, making users objectively more inclined to engage with material aligning with their existing views. Consequently, it reduces demand for diverse perspectives and accelerates the formation of information cocoons. As users spend excessive time in AI-powered recommendation systems, they inadvertently limit exposure to opposing viewpoints or information that doesnt match their personal settings – data deemed irrelevant or untrustworthy.

During unsupervised pre-training phases, Generative Adversarial Networks (GANs) inevitably learn erroneous and biased content from large, complex corpora. These biases may gradually be amplified during dissemination, leading to neglect and discrimination against marginalized groups, ultimately affecting the construction and spread of mainstream ideologies. Particularly when handling decisions involving sensitive information related to gender and religion, GANs might unintentionally exhibit stereotypical perceptions, resulting in unfair judgments that reinforce public prejudices. External factors such as capital influence or political stances also significantly impact GANs content dissemination. Improper capital involvement could turn GANs into profit-driven tools, while political positions may affect algorithms objectives or training processes. Variations in language training datasets may cause GANs to demonstrate bias when addressing certain international disputes, which can later escalate into cultural infiltration and propaganda tools.

In January 2023, the U.S. Department of Justice and Department of Housing and Urban Development filed a joint interest statement regarding a lawsuit alleging that SafeRents algorithmic screening system discriminates against Black tenants, demonstrating that algorithmic bias is becoming a tangible reality. UNESCO Director-General Audrey Azoulay emphasized: "As more people utilize large language models in their work, studies, and daily lives, these new AI tools wield the power to subtly reshape human perceptions. Consequently, even the slightest gender bias in generated content could significantly exacerbate inequalities in the real world." [80, 81]

GAI utilizes adversarial network generation models to produce fake news that closely mimics authentic information. This undoubtedly exacerbates the proliferation of misinformation, undermines credibility and trustworthiness, negatively impacts public decision-making and judgment, and provides opportunities for criminals to spread rumors and manipulate public opinion. The widespread dissemination of false content overwhelms genuine information, making it difficult for people to access accurate and reliable data, ultimately leading to information overload and information chaos. [82]

When GAI combines with collaborative filtering and ranking algorithms, it may trigger "emotional contagion" that intensifies public sentiment toward specific topics. During global election years, the misinformation generated by GAI can reshape information dynamics, infiltrate voters information-seeking behaviors, and ultimately influence election outcomes. [83]

The emergence of Generative Adversarial Intelligence

(GAI) has enhanced the sophistication of information weapons in cognitive warfare, expanded their operational pathways, and enriched its strategic concepts. By controlling GAIs training data sources—such as using datasets with specific political leanings or intentionally adjusting value judgments during manual training phases—we can impose predetermined ideological influences on GAI during its development stage. This process shapes the content generated by GAI in practical applications, effectively serving cognitive warfare objectives. Although OpenAI explicitly prohibited deploying its large language models for military or war-related purposes, its revised statement limiting usage to "not harming oneself or others through products" subtly suggests that under certain circumstances, GAIs application scope may expand accordingly. [84]

2.2.4. Information Leakage: Illegal Acquisition of Personal Information

Privacy violations and data breaches remain persistent threats. With the widespread adoption of digital technologies, critical issues such as data acquisition methods, processing protocols, and the implementation of informed consent mechanisms have become paramount concerns for privacy protection. The extensive use of generative AI has further exacerbated these security challenges, posing significant ethical risks of personal information leakage. The development of generative AI systems requires training on massive datasets to build knowledge bases and corpora, enabling interactive models capable of natural language understanding and content generation. During this process, generative AI not only rapidly scours the internet through web crawling techniques but also utilizes user input as data sources. Training datasets containing sensitive information—including personally identifiable details, images, and voice recordings—often lack clear boundaries in their legitimacy during collection and usage. These models tend to learn and retain such data, which may be disclosed during user interactions, potentially leading to the leakage and proliferation of private information. This creates increasingly severe risks of privacy infringement and data security breaches, with potential consequences including "doxing attacks" and malicious cyber intrusions. Security risks such as extortion." [57]

In April 2023, Canadas Privacy Commissioners Office announced it was launching an investigation into ChatGPTs parent company OpenAI, saying it was involved in "complaints alleging that OpenAI collected, used and disclosed personal information without consent." [85]

In June 2023, a group of plaintiffs filed a class action lawsuit against the creators and distributors of GAIs ChatGPT tool, claiming that OpenAI had captured, collected and processed their data without the consent or authorization of millions of people. [86]

GAI possesses robust capabilities for fragmented information integration and analysis at the output end. It aggregates fragmented personal data from users, conducts cross-referencing analyses, and deeply mines sensitive information and private content. For instance, by analyzing users digital footprints, it extracts hidden sensitive details. Under the "long-tail effect," this poses significant risks to individual rights. Moreover, due to the high costs of model iteration and training, leaked sensitive information cannot be promptly removed, allowing these risks to persist long-term. Such persistent threats ultimately undermine personal peace of mind and dignity. [87]

The collection of such information primarily relies on user input commands, yet the informed consent principle fails to function effectively throughout this process. On one hand, GAIs service terms and privacy policies are lengthy and complex, filled with technical jargon that users find difficult to comprehend. They often habitually click the "Agree" button without fully understanding the scope of permissions granted or their consequences. On the other hand, even when some users attempt to carefully read and understand the clauses, GAIs complexity and technical nature make it challenging for them to accurately assess how personal information will be collected, used, and shared. The lack of transparency prevents users from effectively exercising their privacy rights and right to information, ultimately resulting in a loss of effective control over their personal data. [88]

In the digital age, personal information leaks can lead to severe consequences including identity theft, financial fraud, and even damage to ones reputation. In March 2023, a ChatGPT glitch allowed some users to access others chat history, with even active users being exposed to sensitive details like credit card last four digits, expiration dates, names, email addresses, and payment addresses. Such errors could potentially result in 1. Information about payments is visible to 2% of ChatGPT Plus subscribers. Based on this, Italys privacy regulator believes ChatGPT is suspected of illegally processing personal information in serious violation of the EUs General Data Protection Regulation. [89]

In July 2023, South Koreas Personal Information Protection Commission fined ChatGPT operator OpenAI 3.6 million won for leaking personal information of 687 South Korean citizens. [90]

Given the "black box" nature of deep learning technology, which makes it difficult to effectively implement the purpose limitation principle and transparency principles, GAI may exceed its original information processing objectives and take actions beyond users reasonable expectations and consent. Meanwhile, OpenAI has yet to provide an effective mechanism for individual users to access and inspect personal information stored in their data repositories. This partially undermines users self-determination rights over personal data, making it challenging for them to control the flow of their own information.

2.2.5. Digital and Intellectual Divide and Capital Monopoly: Generation of AI Expands and Intensifies the Imbalance in Development

Amid accelerating technological innovation, expanding technological influence, and increasingly intimate human-machine interactions, the sustained development of Global Artificial Intelligence (GAI) has led to a new digital divide characterized by rapid emergence, profound depth, and wide scope. This disparity now surpasses the ability to bridge existing gaps. Currently evolving into a digital-intelligence divide, its essence lies in a "knowledge gap" or "information gap," manifested through uneven resource distribution, differing application capabilities, and divergent development opportunities across various dimensions. [91]

Under the background of GAI, the digital poverty gap is further widening, and the discourse power of regions with weak science and technology is ignored. The underlying risk is that the values of developed economies are widely spread, which limits and destroys the effective voice of different opinions and diverse ideas, and makes them face the hidden risk of being further marginalized.

The manifestations can be summarized as follows: First, the national intelligence gap, where the economic and social benefits of GAI are predominantly concentrated in the global North. The divide between developed countries and resource-constrained nations will continue to widen. Second, the regional intelligence gap, where structural limitations in core elements like infrastructure, algorithms, computing power, and data deprive underdeveloped regions of crucial developmental momentum. Third, the urban-rural intelligence gap, where socioeconomic elites demonstrate greater willingness and capacity to adopt GAI. Fourth, the intergenerational intelligence gap, where older populations show relatively lower acceptance and proficiency with GAI compared to younger generations who adapt more readily. Significant disparities exist among different groups in terms of GAI awareness, accessibility, usability, and evaluation capabilities. Technological inequality intertwines with existing social inequalities, transforming into disparities in economic income, political participation, and social capital.

Capital-driven Entrenchment: The Formation of an Oligopoly Landscape. The rapid expansion of General Artificial Intelligence (GAI) has introduced new players—universally large tech conglomerates. These corporations leverage extensive user bases and data access to amplify GAI capabilities through economies of scale, thereby strengthening their market dominance. Established firms erect barriers to entry through patents, technical standards, and accumulated data assets, creating formidable obstacles for newcomers. While capital fuels technological innovation and market expansion, it concentrates resources in the hands of a few deep-pocketed players. Moreover, GAIs technical complexity and high barriers to entry have evolved into entrenched technical barriers. This convergence of technological and financial monopolies has forged an oligopolistic market dominated by select enterprises, paradoxically stifling competition and technological progress. In March 2024, Microsoft filed a complaint with EU antitrust regulators alleging that Google maintains a significant competitive advantage in GAI through its vast data resources and AI-optimized chip technology. [92]

In June 2023, the Competition Bureau of the Federal Trade Commission pointed out that GAI relies on a series of necessary inputs and may be concentrated in a few companies, which may lead to data monopolies and affect the ecology of the open source community. [93]

Driven by a handful of giant tech companies, the "AI superpowers" have a significant influence on global AI standards and norms, and this concentration of power will infinitely exacerbate global inequality.

3. Analysis of Legal Governance Strategies for Ethical Risks of Generative AI

"The real core of risk awareness is not now, but in the future," says German scholar Beck. [94]

Profound contemplation and prudent handling of ethical issues in AI education not only guide the future direction of educational AI, but also concern the developmental prospects of human society as a whole. Through coordinated measures including technological mediation, digital ethics promotion, responsibility distribution, and institutional refinement, we aim to resolve ethical constraints in AI applications within education. This approach will better serve the sustainable development of AI technology and facilitate the digital transformation of education.

Currently, generative artificial intelligence is accelerating its penetration into industries and social life. The dual nature of this technological innovation manifests in both emerging opportunities and latent risks. "Modern technology is triggering a global historical transformation in human conditions, one that transcends the specific intentions of its users.... When technological impacts exceed their creators original purposes, they can no longer be considered neutral instruments." [95]

3.1. Technical Mediation: The Coordination and Complementarity Between the "Use" of Technology and the "Way"

The theory of technology mediation was proposed by Dutch scholar Peter-Paul Verveldt. It posits that technology not only shapes peoples perception of the external world but also actively participates in shaping their moral choices and behavioral patterns. Moreover, the mediating role of technology is non-neutral, influencing ethical decision-making processes to some extent. [96]

This provides an effective principle for overcoming the "technological supremacy" and "materialism" in the field of digital justice, and realizing the bidirectional coupling between the usable dimension and the lovely dimension of the application of artificial intelligence in the field of digital justice.

The fundamental ethical risk posed by current AI applications stems from the usurpation of value rationality by instrumental rationality, which triggers a crisis of human subjectivity. The concept of "technological awakening" refers to humanitys cognitive transition from ignorance to clarity in understanding technology, evolving from confusion and ambiguity to clear comprehension. [97]

Humans need to promote the simultaneous governance of algorithms and content:

Algorithm governance is the core foundation to ensure the safe, reliable and responsible operation of GAI. The key lies in improving the transparency and interpretability of algorithms, and ensuring that the design and operation of algorithms meet the standards of law, ethics and social responsibility.

First, the law should clarify GAIs liability chain by establishing developers "algorithmic design responsibility," requiring them to strictly adhere to principles of fairness, impartiality, and transparency during algorithm development. This ensures algorithms do not generate discriminatory or harmful consequences. Second, operators must be held accountable for operational oversight by establishing effective monitoring mechanisms to conduct real-time supervision of algorithmic operations, promptly identifying and correcting deviations or misuse. Third, users should be legally obligated to operate algorithms within compliant frameworks, ultimately preventing data abuse and algorithmic technologies from infringing on legitimate rights.

Secondly, it is crucial to establish rigorous algorithmic review and evaluation mechanisms that assess the impact of data protection measures on critical algorithms, ensuring compliance with legal requirements such as privacy protection. The transparency principle requires developers to provide detailed technical specifications during algorithm design, enabling external reviewers to understand and evaluate decision-making processes while establishing effective feedback and correction mechanisms. The explainability principle mandates developers to clearly articulate the rationale and operational logic behind algorithmic decisions, thereby ensuring traceability and auditability throughout the decision-making process.

Finally, it is essential to refine the tiered protection mechanism for data classification. This requires categorizing data based on its potential impact on national security, public interests, and individual rights. For instance, distinguishing between general data, important data, and core data would standardize usage rules for different categories during algorithm training and applications, thereby ensuring data security and facilitating lawful circulation.

3.2. Digital Goodwill: Service Quality Improvement

Technology has the duality of natural attribute and social attribute. Natural attribute determines the inevitability and irreversibility of technological alienation, while social attribute determines the artificiality and controllability of technological alienation. [98]

The self-organization of technology determines the inevitable result of technological alienation. Although the one-sidedness and limitation of human understanding will lead to the deviation in the understanding and application of technology, it can improve the quality of technical service and guide the intelligent technology towards good governance artificially.

To mitigate ethical risks posed by General Artificial Intelligence (GAI), humanity must prioritize data quality enhancement. As the cornerstone and core element of artificial intelligence technology, big data forms the foundation for all virtual and physical entities in intelligent societies. The accuracy, security, and stability of big data directly determine the proper functioning of smart systems. Therefore, improving data quality serves as a fundamental approach to governance at its source.

First: Data classification and classification. According to the requirements of the Data Security Law, citizens personal privacy protection and application needs require different levels of security protection measures, so as to effectively prevent data leakage and illegal use.

Secondly, it is crucial to verify data sources. By proactively filtering out erroneous information, we can prevent the use of unverified or ambiguous data, thereby ensuring the security and reliability of data sources. Before collecting sample data, clearly inform participants about the collection methods, purposes, and regulations. Strictly prohibit coercing or inducing consent through bundled permissions or default authorization mechanisms.

Thirdly, databases requiring timely updates and optimization. For instance, ChatGPT only incorporated data from 2021, resulting in insufficient accuracy and comprehensiveness in its generated content. Therefore, humans must adopt technological solutions to expand the training scale of data promptly, ensuring both diversity and comprehensive coverage.

Content security norms should be improved. German scholar Marcuse pointed out: "Technical rationality, a form of bounded rationality, has become a new form and force of oppression over people, and has become an ideology." [99]

Technology inherently possesses political and ideological attributes, with AI-generated content being the most direct manifestation of these characteristics. To address this, three key measures should be implemented: First, upholding core values. Intelligent education applications must incorporate ethical filtering mechanisms. The "stop" command not only enables preemptive intervention in sensitive or value-sensitive content but also prevents harmful or illegal information from being published, ensuring generated content aligns with mainstream social values and guides users toward proper political orientation. Second, strengthening identity verification. Rigorous authentication ensures traceability, security, compliance, and credibility of AI-generated content.

When illegal or non-compliant content emerges, it enables swift identification of content creators for accountability. Additionally, implementing algorithmic anti-addiction systems in educational applications—using algorithms to restrict or regulate technology usage duration and recommendation content based on users age, preferences, and social connections—helps prevent excessive immersion and resolves issues like "information cocoons" and "data silos." Third, enhancing content labeling. Clearly identifying AI-generated texts, images, and audiovisual materials allows users to distinguish their sources and nature, thereby increasing transparency and enabling... Strong trust in intelligent education applications.

Principle of informed consent and principle of limitation: In order to protect personal privacy and rights and interests, GAI shall implement the principle of informed consent when processing personal information.

First, enhance transparency. Establish a transparent mechanism spanning the entire lifecycle of large models, covering providers, deployers, and users. This aligns with the requirements for technical understandability and accountability in scientific ethics. Providers and deployers of high-risk applications must disclose critical model information, while entities handling medium-risk applications should fulfill obligations such as providing necessary notifications and obtaining consent, ensuring upstream entities can offer essential technical support to downstream users. [100]

Second, it is essential to ensure clients fully understand the scope of information and the form of consent. Before collecting or using personal data, users must be clearly informed about the intended use, storage methods, and potential risks and consequences. Furthermore, consent must be obtained only when users have given their voluntary and informed approval.

Third, a tiered consent framework should be established. Different categories of personal information have varying impacts on users rights and interests, necessitating a multi-tiered consent system with opt-out mechanisms. For sensitive data directly involving core privacy or capable of precise identity verification, users must provide explicit, separate, and in-depth consent after receiving detailed explanations about usage methods and potential risks. However, for general data, a relatively broad yet clear basic consent approach remains appropriate.

Fourth, comprehensive dynamic control throughout the entire process should be implemented. This requires fully disclosing the purpose, methods, and scope of information usage to users before data collection, ensuring user rights are protected during processing, and promptly implementing corrective measures when issues arise. Given GAIs complex decision-making processes and extensive data utilization, the informed consent mechanism must demonstrate dynamic adaptability – allowing for appropriate adjustments or relaxation of consent requirements under specific circumstances.

The principle of limitation serves as a cornerstone for GAI (General Artificial Intelligence) compliance and ethical governance. This fundamental requirement mandates that when designing and deploying AI systems, humans must establish clear legal boundaries for GAI functionalities, data collection processes, and output outcomes to prevent technological misuse and potential legal risks. Specifically, it requires implementing strict restrictions on the data used by GAI to avoid excessive gathering and purposeless utilization. Furthermore, data collection must adhere to the minimization principle – collecting only the minimum necessary data required to fulfill specific functions, with such data being promptly deleted or anonymized after objectives are achieved, thereby ensuring data security.

3.3. Man-Machine Collaboration: Construction and Remodeling of Multiple Principal Responsibilities

The real era of artificial intelligence must be based on multi-level and all-domain "man-machine collaboration". [101]

For regulators, it is necessary to improve the risk early warning ability in the field of digital justice, ensure that intelligent technology complies with the law of justice, and have the obligation of safety supervision and guarantee.

Firstly, establishing a digital risk early-warning system is crucial. China needs to categorize ethical risks by their types and severity levels while developing corresponding risk management strategies. The EUs Artificial Intelligence Act provides valuable insights for Chinas layered governance framework in addressing ethical risks within educational systems, particularly through its risk classification methodology and regulatory approach.

Secondly, it is crucial to enhance security monitoring and analysis capabilities for smart platforms. By leveraging tools such as Security Insight and Exposition (SIEM), Network Security Situation Awareness (NSA), and Unified Threat Management (UTM), combined with big data desensitization technologies and access control systems, organizations can promptly address cybersecurity incidents and abnormal activity.

Third: Responsible innovation needs to be strengthened. Different subjects should be mobilized to participate in the initiative and enthusiasm of supervision, and decentralized and differentiated regulatory models should be adopted to promote the supervision and innovation of algorithmic applications, so as to realize the benign operation of artificial intelligence and digital judicial applications.

Moderate Governance of AI Applications: In terms of governance approaches, we must first return to legal principles and firmly adhere to the fundamental approach of ensuring rights through power, constraining power with rights, and balancing power through rights. This requires integrating ethical primacy, transparency, fairness, and sustainable human development into technological development to ensure that technological progress benefits humanity without com-

promising human autonomy. Technological research and development must proceed in parallel with ethical constraints, ensuring this powerful technology serves social welfare rather than merely pursuing commercial interests or technological advancement. Mechanisms should be refined to regulate algorithmic rights through public authority, guaranteeing they serve legal principles and rule-of-law order. The boundaries of algorithmic rights must be clarified to ensure their lawful exercise without exceeding their intended functions and purposes. For instance, Japans Ministry of Education allows limited use of GAI tools like ChatGPT in education, reflecting the governments control over technology application boundaries. [102]

The primary task of law is to clarify the boundaries of "power" of technology, which not only includes the definition of the source and scope of exercise of GAI power, but also should strengthen the "proportionality principle" in technology development, that is, technological intervention and decision-making should be moderate and necessary and not beyond the purpose of design.

Firstly, legislation must clearly define the boundaries of government-activated intelligence (GAI) intervention, ensuring technological applications remain within the scope required to achieve established objectives. Technologies should not exceed their original design purposes, avoiding disproportionate interference with individual privacy, freedoms, and social order. For instance, in sensitive fields like healthcare, education, and military applications, their use should be limited to supporting decision-making rather than replacing human judgment. This approach ensures such technologies do not impose unnecessary constraints or excessive impacts on individuals without proper authorization.

Secondly, the law should establish a proportionality review mechanism requiring GAI to undergo rigorous evaluation and examination before deployment in specific applications. This mechanism would assess potential threats and negative impacts on society, the environment, and individual freedoms. Concurrently, regulatory authorities must develop forward-looking yet practical policies and regulations to ensure that technological design and implementation strictly adhere to ethical standards and best practices.

Thirdly, the law should establish industry self-regulation mechanisms that require developers to fully consider the principle of proportionality when designing and deploying General Artificial Intelligence (GAI), while actively assuming ethical responsibilities throughout implementation. These obligations should be reflected in technical specifications, ethical guidelines, risk assessment reports, and other aspects, ensuring developers proactively prevent potential improper interventions during technological development processes.

Fourthly, given the potential ethical risks of General Artificial Intelligence (GAI) technology, legal frameworks should establish emergency response and risk prevention mechanisms to swiftly intervene in applications that may violate the principle of proportionality or cause severe social consequences. This mechanism should span the entire lifecycle of GAI technology—from ethical reviews during R&D stages through continuous monitoring of implementation to post-implementation social impact assessments—thereby forming a comprehensive legal risk management system.

3.4. System Improvement: Standardized Development of Artificial Intelligence Technology System

The refinement of legal frameworks can effectively address the limitations in implementing AI ethical principles. Guided by ethical standards, research findings should be transformed into legislation through legislative processes, establishing laws and regulations as both consensus-building tools and binding benchmarks. These legal provisions define boundaries for GAI applications, thereby fostering parallel development of technological logic and digital judicial systems.

To strengthen institutional support for artificial intelligence (AI) in digital judicial systems, specialized legislation has been enacted. The Interim Measures for the Administration of Generative AI Services introduced in 2023 serve as a landmark regulation promoting AIs healthy development and standardized application. The State Councils inclusion of the AI Law in its 2023 legislative agenda marked Chinas first comprehensive legal framework for AI governance. This provides practical reference points and coordination channels for establishing AI-related legislation in digital judicial applications. Relevant authorities should prioritize legal interpretation to regulate AI application risks within existing frameworks, ensuring legal stability and continuity. Priority should be given to enhancing data security protocols through judicial authorities setting technical standards, implementing corporate guidelines for data collection and usage by enterprises, professionals, and law enforcement agencies, and establishing real-time monitoring mechanisms to address risks throughout data collection, annotation, training, and verification processes. Additionally, improving digital literacy programs will better equip society to adapt to intelligent technologies.

We will promote ethical governance norms for artificial intelligence and make them more effective.

Firstly, ethical principles require optimization to enhance the practical effectiveness of governance guidance. Over a hundred domestic and international guidelines on AI ethics and governance have established macro-level ethical principles for AI development and educational technology advancement. For instance, Chinas "Guidelines for Standardized Ethics Governance in Artificial Intelligence" and other regulations collectively emphasize core values including "shared responsibility," "human welfare," "rights protection," "fairness and justice," "risk prevention," and "transparency." However, improving the validity of these ethical principles and their practical application demands dual efforts: On one hand, we must clarify the definitions of AI ethical principles.

Ambiguous ethical frameworks like the "human welfare" principle could lead to governance challenges—what specific groups does it protect? Might conflicts arise between different groups? Therefore, enhancing the effectiveness of AI ethics governance requires first clarifying its value propositions. On the other hand, we need to refine the relevance and operability of these principles. Overly broad ethical guidelines risk diminishing practical impact, while AI ethical principles must be tailored for concrete applications in digital judicial scenarios.

Secondly, conducting social experiments on artificial intelligence is crucial to provide scientific evidence for ethical governance. These experiments aim to pilot technological pathways through controlled trials, allowing us to clearly understand how AI operates while assessing algorithms potential in building social systems – ultimately enabling better adaptation to intelligent society development. The significance lies in reshaping individual cognition and enhancing understanding of complex social interactions by intervening in human behavior. Current AI social experiments face inherent risks from technological factors and systemic challenges. To address this, leveraging expert expertise is essential to establish collaborative governance mechanisms among stakeholders, thereby promoting systematic implementation of AI social experimentation.

The rapid advancement of data information technology has transformed artificial intelligence into a powerful autonomous entity, providing substantial momentum for expanding practical activities in human society. Within the context of human-machine integration, objective factual chains, personalized value chains, and universal responsibility chains intertwine between humans and intelligent systems, making people and machines mutually measurable benchmarks. [103]

The ancient Greek philosopher Aristotle wrote in the Nicomachean Ethics: "All arts, all disciplines, and likewise, every action and every plan, seem to have their ends in some kind of good, so that good is clearly expressed in the purposes sought by all things." [104]

3.5. Legal Supervision: Coordinate the Positioning and Build a Diversified Security System

Clarifying the instrumental nature of generative artificial intelligence (GAI): While GAI demonstrates remarkable potential in language comprehension, data analysis, text generation, and simulated dialogue—even passing the Turing test—thereby blurring traditional human-machine boundaries in certain scenarios, it remains fundamentally a product of human intelligence. As a tool created by humans and ultimately designed for human use, its core purpose lies in assisting rather than replacing humanity. GAI does not possess autonomous moral judgment or ethical decision-making capabilities. The technologicals ethical standing exists neither as an inherent attribute of the technology itself nor as a human-imposed characteristic, but rather resides within the dynamic relationship between humans and technology. [105]

The fundamental mission of law lies in maintaining social order and justice while regulating human behavior, rather than directly controlling technological tools themselves. The actions of General Artificial Intelligence (GAI) are entirely determined by user commands, goal-setting mechanisms, and algorithmic design. The corresponding ethical responsibilities should rightfully fall on decision-makers—including designers, developers, operators, and users—whose rationality enables informed choices. In the AI era, our understanding of GAI demands fundamental rethinking: When developing intelligent models, we must establish "good" and "bad" standards centered on human interests; During industrial development, prioritizing "algorithmic safety" while strengthening risk constraints; When establishing human-machine symbiosis frameworks, delineating artificial intelligences "living space" and "rights domain" to safeguard humanitys dominant position in power dynamics, ensuring its development and application align with human values and ethical principles. [106]

Firstly, legal frameworks must ensure controllability. Functional safety regulations explicitly require GAI developers to incorporate Human-in-the-Loop (HITL) mechanisms during the design phase, enabling timely intervention when GAI systems malfunction or perform hazardous tasks. Additionally, legislators should define accountability thresholds for GAI systems. Specifically, when system operations exceed safe parameters, clear legal responsibilities must be established for manufacturers, operators, and users. This ensures traceability of liability to specific entities in cases of misuse or safety incidents.

Second: the law needs to be clearly interpretable. In the process of legislation, it needs to be embedded in the form of legal provisions.

First, interpretative principles should be established to define the scope of relevant concepts. Meanwhile, in terms of regulatory models, necessary interpretative obligations should be imposed on high-risk AI developers and providers. Finally, in terms of professional models, supportive regulation should be prioritized while providing support for highly interpretable models.

Third: The law needs to be clear and predictable. Legislators need to design embedded risk assessment and management.

The framework requires developers to conduct detailed prospective analysis and regular transparent audit reports on the potential risks of GAI, as well as to disclose the source of training data, training process and decision-making mechanism of their models, so that regulators and the public can monitor and evaluate the behavior of the system.

Fourth: The law needs to clarify the usability. On the premise of adhering to the principle of moderate regulation of technology, the applicable scope and data use standards of GAI must be clearly defined in the form of regulations, and

the usage scenarios, professional requirements and responsibilities of GAI in different fields should be defined.

To ensure that GAI is open, inclusive and equitable to bridge the digital divide created by GAIs rapid development, international cooperation, national policies and collaboration from all sectors of society are needed.

Firstly, at the international level, the global community must establish specialized international treaties or multilateral agreements to define standards for technical assistance and knowledge transfer. This ensures developing countries can access and utilize GAI technologies equitably while reducing technological hegemony and knowledge barriers. Simultaneously, the legal framework for international cooperation should include efforts to establish global ethical norms for AI development, preventing the exacerbation of social inequality and technological dominance during collaborative advancements in GAI technologies. Through cross-border project partnerships, technical assistance, and educational training programs, developing nations can enhance their R&D capabilities and application proficiency in GAI fields. Ultimately, this approach aims to bridge the persistent "AI divide" between developed and developing countries.

Secondly, at the national level, governments should implement inclusive and prudent policies. This includes increasing legislative investments in digital infrastructure development—particularly in network coverage, device deployment, and data infrastructure—to ensure equitable access to digital resources across all regions. For instance, governments could enact Digital Infrastructure Laws requiring coordinated efforts between enterprises and local authorities, thereby accelerating the adoption of internet services and digital devices in remote areas.

Third, the government should prevent a few technology giants from monopolizing GAI through administrative supervision, so as to ensure that small and medium-sized enterprises and innovative companies can actively participate in technological development under the environment of equal competition.

The content generated by AI significantly shapes users self-perception and the social information ecosystem. Content governance aims to ensure that generated content remains authentic, accurate, beneficial, and harmless. This urgently requires establishing a comprehensive content review and feedback system for real-time monitoring and management, enabling timely correction of errors and inappropriate information. The governance framework should clarify responsibilities between upstream model developers and downstream application deployers. Upstream developers must maintain overall system security through compliance obligations in data collection/processing, algorithm design/training, and model optimization/updating. Downstream deployers are responsible for ensuring generated content complies with legal, ethical, and regulatory requirements, including conducting content reviews to prevent improper dissemination and meeting industry-specific compliance standards. To enhance governance effectiveness, a "crowdsourced governance" mechanism can facilitate collaboration between stakeholders. This approach enables joint content management responsibility fulfillment while meeting minimum legal requirements. [107]

To effectively assess and ensure compliance of generated content, a multi-dimensional quality evaluation system should be established, covering key metrics such as accuracy, authenticity, relevance, originality, and acceptability. To safeguard the fairness and impartiality of AI-generated content, laws should mandate regular fairness reviews of GAI systems to ensure their output contains no discriminatory, biased, or exclusionary elements.

The high autonomy of AI technologies and the unpredictability of generated content urgently require a more dynamic and diversified legal framework to regulate emerging risks and ensure technological innovation stays aligned with public interests. To overcome these limitations, legislation must introduce multi-stakeholder collaboration, supervision, and checks-and-balances mechanisms. This approach should foster positive interactions among government, market, enterprises, and consumers while adhering to objective and fair legal standards for effective governance of GAI. Such collaborative governance models help establish an inclusive yet prudent regulatory framework that promotes long-term technological advancement while ensuring comprehensive oversight. As direct operators of GAI and creators of risks, platforms bear responsibility for ensuring the safety, transparency, and fairness of their technological applications. Within legal boundaries, platforms must implement necessary technical measures and management protocols to ensure content compliance with social order and legal norms. For instance, establishing internal compliance systems—including strict data processing guidelines, algorithm transparency requirements, and content moderation standards—through self-review mechanisms can prevent and mitigate risks. Additionally, developing technical usage rules and emergency response measures will strengthen corporate accountability and platform governance. The regulation and restraint of the platform can effectively prevent data leakage, false content generation and other chaos. [108]

Admittedly, platform censorship mechanisms have inherent limitations. They may overlook risk prevention due to commercial interests or fail to comprehensively identify and manage risks owing to technical expertise constraints. Therefore, there is an urgent need to establish a multi-stakeholder governance framework where governments, industry organizations, civil society, and other entities collaborate in regulating Global Artificial Intelligence (GAI). [109]

As the primary regulatory authority, the government should establish a clear and forward-looking legal framework by creating specialized regulatory bodies to monitor platform review practices in real time. This will ultimately define the legal responsibilities of platforms during the review process to

standardize their operations. By implementing a regulatory sandbox mechanism, we can effectively control technological risks while preserving innovation, thereby providing a new and feasible governance approach for Global Artificial Intelligence (GAI). The regulatory sandbox could set unified entry thresholds and restrictive conditions to ensure fair treatment and limited exemptions for all GAI enterprises. It should encourage innovative approaches and tolerate trial-and-error processes to validate the scientific validity and feasibility of GAI service models, application mechanisms, and operational workflows.

Sandbox environments typically employ virtualization technology to simulate operating systems and hardware configurations, enabling regulators to adapt oversight frameworks and implement differentiated approaches for GAI products with distinct risk profiles. This methodology maintains necessary flexibility for innovators within regulatory parameters, fostering constructive collaboration between regulators and GAI innovators. Industry associations should serve as vital intermediaries by establishing industry standards, self-regulatory protocols, and facilitating knowledge-sharing and best practice exchanges. Meanwhile, civil society can enhance public awareness of GAI risks through mechanisms like public oversight and social advocacy campaigns.

The active participation of the general public in legal governance helps build a shared value consensus across society, thereby fostering a healthy ideological and public opinion environment. This ultimately ensures that platform review mechanisms align with both public expectations and societal ethical standards. From this perspective, only by combining platform oversight with multi-stakeholder co-governance can we fully leverage each partys strengths to form a regulatory synergy. This approach enables comprehensive and effective risk management while achieving a balanced coordination between technological innovation and social security alongside public interests.

Traditional regulatory frameworks such as antitrust laws primarily focus on corporate market share and competitive practices. However, Global Artificial Intelligence (GAI) demonstrates exceptional complexity in practical applications, requiring cross-domain technical collaboration, data sharing, and global operations. Consequently, conventional market concentration analysis standards have become inadequate for addressing GAIs unique characteristics. Antitrust legislation must therefore not only monitor market consolidation but also strengthen oversight over centralized control of technologies, data, and algorithms to ensure fair competition in the marketplace.

Undoubtedly, the dominance of GAI technologies is fundamentally tied to platform economics. These enterprises typically operate within large-scale platform architectures that create powerful network effects through integrated services like content generation, recommendation engines, and advertising systems. This enables a few dominant platforms to monopolize market share, ultimately exacerbating asymmetry in competition. Furthermore, the platform economy inherently features network effects, user lock-in, and data lock-in mechanisms, resulting in competitive structures distinct from traditional industries. Therefore, Chinas Anti-Monopoly Law could adopt legislative principles from the EUs Digital Markets Act to establish specialized regulations for the platform economy. Such legislation should prohibit improper self-preferential practices by platforms, specifically banning deliberate favoritism toward their own products/services within the same ecosystem while unfairly treating competitors. This approach would effectively curb monopolistic behaviors and ensure fair market competition.

From a macro perspective, antitrust regulatory mechanisms should maintain technological neutrality to prevent unfair privileges and restrictions on specific technologies or enterprises. By encouraging data sharing and openness while establishing reasonable policies and standards for data exchange, we can reduce the difficulty and cost for small and medium-sized enterprises (SMEs) in accessing critical data, ultimately lowering market entry barriers.

The anti-monopoly law should also strengthen the governance of technology application through industry regulation, etc. The cooperating regulatory agencies need sufficient response capacity and cooperation to deal with the competition problems that GAI may cause in different fields and industries.

The Anti-Monopoly Law should strengthen oversight of capital operations to prevent large enterprises from achieving technological and market monopolies through capital concentration. From a micro perspective, antitrust enforcement agencies should be granted model accountability rights, requiring companies to transparently explain the decision-making logic behind GAI models. Platform enterprises must also conduct regular reviews of critical algorithms and submit detailed audit reports for regulatory bodies and independent third parties to examine and evaluate. Furthermore, the law should explicitly define data exploitation and model preference behaviors as illegal acts of abusing market dominance, with specific legal liabilities and penalties established. The antitrust review mechanism should particularly focus on monitoring GAI enterprises acquisition of technological and market control through capital operations. For large corporations obtaining exclusive licenses for GAI models via investments and acquisitions, such concentration behaviors should be treated as mandatory antitrust filings and reviews.

To safeguard personal privacy and data security: First, regulatory bodies should implement comprehensive oversight of data storage, transmission, usage, and sharing processes while establishing corresponding legal standards to prevent data misuse and illegal transactions. Second, the principle of limitation requires strict definition of GAI application scenarios, ensuring their functions and purposes remain within reasonable and legal boundaries. Laws must establish clear regulatory frameworks to standardize GAI technology usage,

preventing its abuse in creating false information, infringement activities, and political manipulation. Finally, the principle of limitation demands necessary review and control over GAI-generated content. Given potential negative effects like discrimination, bias, or malicious manipulation during content generation, legislation should mandate mandatory review mechanisms—including real-time monitoring and post-event accountability. This ensures developers and users maintain controllable boundaries throughout technological applications, ultimately avoiding ethical and legal risks caused by technology spiraling out of control.

The improvement of GAIs personal information protection compliance system cannot be separated from strengthening the transparency of data processing and the interpretability requirements of algorithms in legislation, so as to strengthen the protection of data subjects rights.

First, GAI must ensure all users can fully comprehend how their personal data is processed and the rationale behind algorithmic decisions. Developers of GAI systems should clearly inform users about data collection, usage, and storage methods, while providing comprehensive explanations of algorithm design and decision-making processes. This is particularly crucial in sensitive scenarios involving automated decisions, sentiment analysis, and facial recognition technologies, where explicit justification for algorithmic choices and detailed analytical reports must be provided.

Second, it is essential to safeguard data subjects rights, particularly in information access, correction, deletion, and objections to processing. Data subjects should have the right to view their own data, correct inaccuracies, delete unnecessary information, or object to automated decisions based on their personal data. To ensure effective exercise of these rights, legal provisions or industry standards should mandate that GAI developers provide simple and transparent channels, enabling data users to query and manage their information anytime, anywhere.

Thirdly, GAIs compliance governance not only requires adherence to relevant laws and regulations but also necessitates a multi-level approach from within the enterprise to establish a comprehensive, dynamic, and interdisciplinary compliance management system. This involves formulating implementation plans for the compliance framework, designing compliant organizational structures, conducting risk assessments, and other critical steps. Consequently, compliance management spans extensive areas involving multiple departments, operations, and personnel. It demands the establishment of a robust, vertically integrated work system to ensure efficient operations and prevent accountability evasion.

In addition to effectively regulating GAI-related behaviors, the law must also require the establishment of feasible implementation mechanisms and regulatory systems to ensure that compliance requirements are fully implemented in practice.

First, it is imperative to establish an independent and authoritative data protection regulatory body that covers areas such as data protection, cybersecurity, and AI ethics, with the capacity to address complex data security issues. The regulatory body should conduct regular reviews of GAI development and applications, particularly focusing on sensitive data usage and algorithmic decision-making in high-risk scenarios.

Secondly, legislation should grant regulatory bodies robust enforcement powers, including but not limited to administrative penalties, economic sanctions, and operational restrictions. Authorities should implement progressive disciplinary measures: starting with warnings and corrective actions, followed by stricter penalties such as fines, business suspensions, or license revocations if non-compliance persists. For enterprises that intentionally conceal data breaches or fail to fulfill their obligations to protect data subjects rights, enhanced deterrence through substantial fines, legal proceedings, and criminal prosecution is crucial. Penalties must be proportionate, with both upper and lower limits established based on the scale and severity of violations to ensure fairness and reasonableness. Finally, international cooperation and global coordination form another critical pillar for strengthening oversight. Given that Global Access Information (GAI) applications are inherently cross-border—where data transmission and processing frequently transcend national boundaries—single-country legal frameworks often prove inadequate in addressing the substantial privacy risks posed by global data flows.

Therefore, countries should strengthen international cooperation on data protection and GAI ethics. Therefore, transnational agreements should be signed and international regulatory coordination mechanisms should be established to share compliance standards and information, so as to reduce regulatory blind spots and legal conflicts.

3.6. Ethics: Strengthening Ethical Norms for the Promotion of Human Well-Being

We should not only understand generative AI as a tool or means to improve life and production efficiency, but also correctly view the huge impact of generative AI on human beings from the perspective of the interaction between technology and people, so as to find feasible countermeasures to prevent the ethical risks brought by technology.

Strengthening ethical norms to enhance human well-being. On September 26, 2021, the China New Generation Artificial Intelligence Governance Committee released the "Ethical Norms for New Generation Artificial Intelligence", proposing fundamental ethical principles such as "enhancing human welfare" and "promoting fairness and justice". These norms provide ethical guidance and value orientation for AI activities, reflecting the ethical regulatory requirements from a public governance perspective. The forefront of technological innovation often becomes the arena where commercial interests are most fiercely pursued. Generative AI innovations may

bring significant commercial benefits to related tech enterprises. Adhering to ethical norms that promote human well-being can establish value criteria for balancing benefits and risks between enterprises and governments. Whether it is AI technology innovators, government science and technology decision-making departments, or the general public using generative AI, all should prioritize enhancing human well-being as the core value driving technological innovation and social application. In *The Superhuman Code*, Morel proposed: Are we building a better future for humanity with the help of grand technologies, or are we constructing a superior technological future at the expense of human welfare? [110]

The key to the problem is to ensure that the priorities we assign to technology and the management methods we give it are in the best interests of all humankind. [111]

In the context of generative AIs transformative yet risky development trajectory, we must uphold the fundamental principle that AI should serve humanity and society during its innovation, deployment, and application. It is crucial to reinforce ethical standards that enhance human welfare, translating these principles into comprehensive ethical practices across all dimensions—from model development and policy responses to social education and international governance.

Enhancing the explainability and transparency of generative AI is crucial in its practical applications. Users place greater trust in content they can better understand, which makes it essential to leverage technological solutions that improve AIs interpretability and openness. By making research outcomes more comprehensible and verifiable, we can effectively mitigate ethical risks such as algorithmic bias, social prejudice, and privacy violations. The European Unions General Data Protection Regulation (GDPR) mandates users right to request explanations when using algorithms, thereby reducing privacy risks stemming from the black-box nature of technological systems. [112]

The Interim Measures for the Administration of Generative Artificial Intelligence Services issued by China also emphasize the need to enhance the transparency of generative AI services, improve the accuracy and reliability of generated content, and provide normative guidance for preventing ethical risks associated with generative AI. [113]

Therefore, it is essential to embed value propositions such as enhancing human welfare and fulfilling public interests into generative AI algorithm design and model construction through coding. Clear design standards for generative models should be established, with continuous interpretability evaluations enabling process optimization and dynamic adjustments. A retrospective mechanism must be implemented for data sources used in large model training to improve the explainability and traceability of generated content. This approach will enhance understanding and control over model behavior, ultimately fostering the development of secure, privacy-protective, transparent, interpretable, equitable, and responsible generative AI systems that prevent potential harm to individuals, businesses, and society. [114]

Emphasis should be placed on the ethical regulation of generative artificial intelligence. The technological innovation iteration cycle of generative AI is continuously shortening, making it imperative to strengthen governance measures at the policy level. In addition to formulating laws and regulations targeting generative AI that clearly define conditions for its development, deployment, and use, as well as provisions regarding system transparency, fairness, accountability, and privacy protection, greater attention should be paid to enhancing ethical regulatory measures, standardizing ethical review processes, balancing the relationship between technological innovation and social responsibility, and promoting the ethical development of generative AI. First, specialized ethical review institutions or committees should be established to oversee the ethical compliance of generative AI projects. These institutions shall examine the design of generative AI systems, regulate data usage, and standardize the collection, storage, and processing of personal data by such systems, ensuring data utilization meets ethical standards and requirements. China implemented the "Measures for Scientific and Technological Ethics Review (Trial)" in December 2023, providing guidelines for regulating scientific research, technological development, and other scientific activities. This initiative helps strengthen the prevention and control of technological ethics risks and promotes responsible innovation. [114]

We should establish a continuous ethical monitoring and policy adjustment mechanism. Given the rapid development of generative AI technology, risk prevention policies must be dynamically updated in adaptive and randomized ways. Therefore, we need to implement ongoing ethical monitoring and review processes to promptly update or refine regulatory frameworks. Simultaneously, its crucial to build a collaborative governance system involving multiple stakeholders in ethical risks related to generative AI. This includes fostering dialogue among ethicists, technologists, policymakers, and other relevant parties. By comprehensively considering technological aspects, legal regulations, ethical standards, and social participation, we can ultimately implement appropriate preventive measures throughout the entire lifecycle of generative AI design, development, and application.

To strengthen the innovation accountability and industry self-regulation of generative AI technology enterprises. First, companies should establish dedicated ethics committees or expert teams to oversee and guide generative AI projects. These committees or teams should include ethicists, legal experts, and technical specialists to ensure ethical training, development, and application of generative AI technologies. Ethical considerations must be embedded throughout the entire process of technological innovation and usage, translating corresponding ethical norms into technical frameworks. This approach transforms generative AI into a tool for achieving moral objectives, making privacy protection through technology and design an indispensable component

of data protection mechanisms. Technologies such as encryption, anonymization, and differential privacy play crucial roles in this process. [115]

We should strengthen ethical training for enterprise technical personnel to enhance their awareness and sensitivity to the ethical risks of generative AI. Training content could include ethical principles, case studies, and decision-making models, thereby embedding practical ethical guidelines into AI product design. Concurrently, process supervision should be implemented to ensure fair use of training data while emphasizing algorithmic transparency and privacy protection. It is crucial to promote industry organizations in establishing self-regulatory mechanisms for formulating ethical standards and criteria for AI innovation. These ethical guidelines should cover data privacy, fairness, transparency, and accountability, guiding and regulating technical personnels conduct in GAI projects while fostering social responsibility awareness. Enterprises should establish comprehensive internal review systems—including project evaluation, risk management, and complaint handling—to identify and resolve potential ethical risks. Finally, companies must maintain transparency by disclosing key information such as GAI project design principles, data usage methods, and algorithmic logic. Strengthening communication and collaboration with stakeholders like governments, academia, and social organizations will facilitate joint discussions on the ethical and social responsibilities of generative AI, ultimately driving industry self-regulation and innovation accountability practicable.

Advance public awareness and education on ethical risks in generative AI. First, leverage both traditional and new media platforms to disseminate knowledge about ethical risks in AI-generated content. Through newspapers, television, and online channels, we should inform the public with relevant information and warnings, "enhancing their ability to identify AI-generated content and cultivate critical thinking skills to address issues of fake news and misleading information." [37]

Secondly, it is essential to fully leverage the role of scientific and technological associations. We should encourage social organizations, non-profit institutions, and NGOs to conduct educational campaigns and awareness programs. By organizing public education workshops and seminars, we can introduce the fundamental principles, application fields, and ethical risks of Global Artificial Intelligence (GAI) to the general public. This approach will attract broader societal participation in disseminating GAI ethical risk knowledge, thereby fostering a collaborative effort among multiple stakeholders.

Thirdly, it is crucial to enhance public discourse on the ethical risks of General Artificial Intelligence (GAI) and their regulatory frameworks. By organizing open hearings, seminars, and other forums, we should establish dedicated ethical risk hotlines and online platforms to provide accessible consultation services. This initiative will help address public concerns and resolve related issues, ultimately increasing societal engagement in shaping GAI ethics regulations while ensuring greater transparency and democratic participation in policy-making processes.

Fourth, integrate GAI ethical risk knowledge into the schools curriculum system and research framework to gradually cultivate students ethical thinking and sense of responsibility. Emphasis should be placed on combining GAI ethical risk research with Chinas social realities, thereby enabling reasonable risk prediction and assessment, constructive risk communication, and risk management tailored to Chinas actual conditions.

We need to establish an international cooperation mechanism for the ethical risk governance of GAI. As GAI spreads and is applied globally, ethical risk governance requires cross-border international cooperation. It is essential to promote countries to jointly formulate and implement transnational ethical standards and norms, including technical standards, ethical guidelines, and legal frameworks to prevent GAI from being abused or used in transnational criminal activities. In recent years, China has successively issued policy documents such as the "Chinas Position Paper on Regulating Military Applications of Artificial Intelligence", the "Chinas Position Paper on Strengthening Ethical Governance of Artificial Intelligence", and the Global Initiative on AI Governance. These documents have proposed fundamental approaches and constructive solutions for AI governance to the world, providing crucial guidance for promoting ethical development of artificial intelligence. In advancing international cooperation on GAI ethical risk governance, we should establish robust mechanisms to facilitate participation from multiple stakeholders in the international community. "Diverse governance entities within the AI socio-technical system possess varying authorities, resources, interests, and constraints. Through extensive interactions via formal and informal channels, they form a composite governance mechanism." By sharing information on GAI and its derivative ethical risks, conducting collaborative research and technological innovation, and strengthening cross-border regulatory cooperation, we can collectively address the challenges of GAI risk governance. Ultimately, this will drive the safe, reliable, and controllable development of GAI develop. [116]

In November 2021, UNESCO adopted the "Recommendation on Ethics in Artificial Intelligence" – a global agreement that has significantly advanced multilateral cooperation in international AI governance. In February 2024, the Second Global Forum on AI Ethics facilitated a joint commitment among participating global tech institutions and companies to "build AI that promotes public welfare". Furthermore, international organizations should actively promote global exchange initiatives and talent development programs in the field of Generative Artificial Intelligence (GAI), working together to ensure the ethical advancement and application of AI technologies for positive societal outcomes. [117]

4. Conclusion

Generative Artificial Intelligence (GAI) has rapidly gained global prominence through its unique content generation capabilities, achieving remarkable breakthroughs across various fields. While scholars and experts are exploring the feasibility of GAI implementation, they simultaneously raise concerns about potential ethical risks and challenges it may pose. As humanity enjoys the technological dividends of GAI, we must adopt an inclusive yet cautious perspective toward its ethical implications. Although GAI currently lacks self-aware superintelligence, its exceptional performance in imitation and creation inadvertently amplifies biases and discrimination inherent in training datasets. This could lead to both potential and actual infringements on individual privacy rights, freedom of expression, and right to information.

Furthermore, the expanding application of science and technology is increasingly concentrating power in the hands of large tech corporations that control vast amounts of data, algorithms, and computing resources, thereby exacerbating social inequality. To address this, GAI must ensure ethical principles, social values, and legal compliance throughout its design, development, deployment, and usage processes. It is crucial to strengthen existing legal frameworks, enhance ethical guidance and social accountability, and prevent technological progress from sacrificing fundamental human rights. Through collaborative efforts and sustained dialogue among stakeholders, we can gradually steer GAI toward a more equitable, transparent, and sustainable development path, ultimately achieving harmonious coexistence between technology and humanity.

The global landscape is witnessing a vigorous development of a new technological revolution and industrial transformation. As General Artificial Intelligence (GAI) undergoes rapid innovation iterations, it continues to empower human society across multiple domains: serving as an essential driving force for innovation and efficiency enhancement in the digital age, while simultaneously raising public ethical concerns due to its inherent uncertainties, risks, and complex disruptive social impacts. Assessing and addressing GAIs ethical risks proves crucial for protecting individual rights, upholding fairness and justice, and promoting social harmony. Therefore, we should approach GAI technological progress with a cautious yet forward-looking ethical perspective, ensuring that GAI truly creates benefits for humanity.

Abbreviations

GAI	General Artificial Intelligence
NGO	Non-Governmental Organization
BNNs	Biological NeuralNetworks
AI	Artificial Intelligence

Author Contributions

Chunlei Sun is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] SEARLE JR. Minds, Brains, and Programs [J]. The Behavioral and Brain Sciences, 1980(3) 417−457.

[2] Li Baiyang, Bai Yun, Zhan Xini, et al. Technical Characteristics and Evolution of Artificial Intelligence Generated Content (AIGC) [J]. Library & Information Science, 2023(1): 66-74.

[3] RIOS-CAMPOS C, LUZURIAGA VITERI J D C, PALMA BATALLA E A, et al. Generative artificial intelligence [J]. South Florida journal of development, 2023, 4(6): 2305-2320.

[4] Qin Tianhong. Global Forum Discusses AI Governance Framework [N]. Economic Information Daily, 2024-02-07 (4).

[5] Meng Lingxiao. Generative Artificial Intelligence Triggers "Three Transformations" in the Computing Power Industry [N]. China Science News, 2023-12-13 (3).

[6] Henry Kissinger, Eric Schmidt, Daniel Huttenlohr. The Age of Artificial Intelligence and the Future of Humanity [M]. Huli Ping and Feng Jun, Translators. Beijing: CITIC Publishing Group, 2023: 19-20.

[7] GOERTZEL B, PENNACHIN C. The Novamente Artificial Intelli- gence Engine [M] //Artificial General Intelligence (Cognitive Technologies). Berlin: Springer Berlin Heidelberg, 2007.

[8] Xiao Yanghua. Generative Language Models and General Artificial Intelligence: Connotation, Path and Inspiration [J]. Peoples Forum Academic Frontier, 2023(14) 49-57.

[9] Li Xueyao. The Legal Nature of Artificial Intelligence Ethics [J]. Chinese and Foreign Law, 2024, 36(4): 884-898.

[10] Bian Chun. Generative AI shows tremendous potential! McKinsey: Could contribute equivalent to one UK economy annually to global GDP [EB/OL]. Cailian Press, June 14, 2023 (October 2, 2024). https://www.clss.cn/detail/1378266

[11] Xue Tao. UN General Assembly passes first global resolution on artificial intelligence [EB/OL]. Xinhua News Agency, (2024-03-22) [2024-10-02]. http://www.news.cn/20240322/b87aab54aeca45b6adf574eea56851c2/c.html

[12] Sun Ninghui. Lecture 10 of the Special Lecture Series of the 14th NPC Standing Committee on the Development of Artificial Intelligence and Intelligent Computing [EB/OL]. China NPC Network, (2024-04-30) [2024-10-02]. http://www.npc.gov.cn/npc/c2/c30834/202404/t20240430_436915.html

[13] THOMSEN K. The Ouroboros Model, Proposal for Self−Organizing General Cognition Substantiated [J]. AI, 2021(1): 89−105.

[14] Bi Wenxuan. The Risk Regulation Dilemma and Resolution of Generative Artificial Intelligence: A Perspective on the Regulation of ChatGPT [J]. Comparative Law Research, 2023(3): 155-172.

[15] Zhang Linghan. Legal Positioning and Hierarchical Governance of Generative Artificial Intelligence [J]. Modern Law, 2023, 45(4): 126-141.

[16] Chen Bing. Legal Basis for the Trustworthy Development of Generative Artificial Intelligence [J]. Journal of Shanghai University of Political Science and Law (Legal Theory Series), 2023, 38(4): 13-27.

[17] Liu Wenjie. Why Transparency and How Transparency: The Construction of Transparency Rules in Artificial Intelligence Law [J]. Comparative Law Research, 2024(2): 120-134.

[18] Feng Zixuan. Ethical Position and Governance of Generative Artificial Intelligence Applications: A Case Study of ChatGPT. Journal of East China University of Political Science and Law, 2024, 27(1): 61-71.

[19] Guo Xiaodong. Risks of Generative Artificial Intelligence and Its Inclusive Legal Governance [J]. Journal of Beijing Institute of Technology (Social Sciences Edition), 2023, 25(6): 93-105.

[20] See Michael Kan. Hacker Deepfakes Employee's Voice in Phone Call to Breach IT Company [EB/OL]. (2023−09−15) [2024−10−16]. https://www.pcmag.com/news/hacker-deepfakes-employees-voice-in-phone-call-to-breach-it-company

[21] See Brandon Vigliarolo If you give Copilot the reins, don't be surprised when it spills your secrets [EB/OL]. (2024−08−08) [2024−10−16]. https://www.Theregister.com/2024/08/08/copilot_black_hat_vulns/

[22] Cai Lin. Logical Justification and Construction Approach of Legal Ethics for Algorithmic Technology [J]. Journal of Legal Science (Journal of Northwest University of Political Science and Law), 2024, 42(4): 92-104.

[23] Yan Kunru. Interpreting Artificial Intelligence: Origins, Approaches, and Practices [J]. Exploration and Controversy, 2022(8): 102-109, 178.

[24] CHALMERS D J. The Conscious Mind: in Search of a Theory of Conscious Experience [M]. New York: OxfordUniversity Press, 1996.

[25] Mitchell Landman. Philosophy of Anthropology [M]. Yan Jia, translator. Guiyang: Guizhou Peoples Publishing House, 2006.

[26] Jiang Penghui. Research on the Neurophilosophy of Brain and Cognition [D]. Taiyuan: Shanxi University, 2023.

[27] Li Jianhui. The Formalization of the Mind and Its Challenges: Philosophy of Cognitive Science [J]. China Philosophy Yearbook, 2018(1): 365-366.

[28] Roger Penrose. New Brain of Emperor: About Computer, Human Brain and Physical Laws [M]. Xu Mingxian and Wu Zhongchao, Translators. Changsha: Hunan Science and Technology Press, 2007.

[29] Qasem F. ChatGPT in scientific and academic research: future fears and reassurances [J]. Library Hi Tech News, 2023, 40(3): 30−32.

[30] George A S, George A S H, Baskar T, etal. The Allure of Artificial Intimacy: Examining the Appeal and Ethics of Using Generative AI for Simulated Relationships [J]. Partners Universal International Innovation Journal, 2023, 1(6): 132−147.

[31] Zhang Yongxue. From Automation Technology to Generative Artificial Intelligence: A Study on Skill Heterogeneity in the Impact of Technology on Workers [J]. Sociology Research, 2024, 39(4): 69-91.

[32] Compiled by the Marx Engels Lenin Stalin Works Compilation Bureau of the CPC Central Committee. Marx and Engels Collected Works (Volume 1) [M]. Beijing: Peoples Publishing House, 2009: 501.

[33] Compiled by the Marx Engels Lenin Stalin Works Compilation Bureau of the CPC Central Committee. Marx and Engels Collected Works (Volume 1) [M]. Beijing: Peoples Publishing House, 2009: 159.

[34] Compiled by the Marx Engels Lenin Stalin Works Compilation Bureau of the CPC Central Committee. Marx Engels Collection (Volume 1) [M]. Beijing: Peoples Publishing House, 2009: 157.

[35] Bao Jin, Huang Jing. Marxist Examination of the Human-Machine Relationship in the Intelligent Era [J]. Ideological and Theoretical Education, 2023(11): 85-91.

[36] Cheng Le. The Current Status, Challenges and Prospects of Generative AI Governance [J]. Peoples Tribune, 2024(2): 76-81.

[37] Xing Can. The launch of the Wen Sheng video model Sora brings both transformation and risks [N]. China City News, 2024-02-26 (4).

[38] Representatives of the 27 EU countries unanimously supported the text of the AI Bill [EB/OL] (2024-02-03) [2024-05-08]. https://baijiahao.baidu.com/s?id=1789848941640499406&wfr=spider&for=pc

[39] Zhang Ting. Ethics and Risk Governance of Artificial Intelligence: An Analysis [J]. Zhongzhou Academic Journal, 2022(1): 114-118.

[40] Luo Huijun, Zha Yunlong. Global Governance Transformation in the Era of Artificial Intelligence and Chinas Response [J]. Journal of Shanghai Jiao Tong University (Philosophy and Social Sciences Edition), 2023(12): 13-24.

[41] Yan Kunru. The Evolution of Research on Technological Risks and Analysis of Its Development Trends [J]. Academic Research, 2017(2): 26-32.

[42] Meng Tianguang and Li Zhenzhen. Governance Algorithm: Ethical Principles of Algorithmic Risks and Their Governance Logic [J]. Academic Forum, 2022(1): 9-20.

[43] Zhi Zhenfeng. Information Content Governance of Generative Artificial Intelligence Large Models [J]. Political and Legal Forum, 2023, 41(4): 34-48.

[44] Warmling D L, Bastone P. Freud e o descentramento da subjetividade: o inconsciente como via de recusa consciencialista [J]. Griot: Revista de Filosofia, 2023, 23(2): 79−98.

[45] Singh O P. Artificial Intelligence and Psychoanalysis: A New Concept of Research Methodology [J] NPRC Journal of Multidisciplinary Research, 2024, 1(1): 19−27.

[46] Bontcheva K, Papadopoulous S, Tsalakanidou F, et al. Generative AI and disinformation: recent advances, challenges, and opportunities [J]. 2024.

[47] Dai Jinping, Tan Yangyang, Ideological Risks and Countermeasures of ChatGPT and Other Generative Artificial Intelligence [J]. Journal of Chongqing University (Social Sciences Edition), 2023, 29(5): 101-110.

[48] Gao Lianghua: "Out of Control Technology and Human Responsibility: On the Frankenstein Problem", Science and Society, No. 3, 2016.

[49] [French] Bernard Stiegler: "Technology and Time: The Fault of Epiemus", translated by Pei Cheng, Nanjing: Yilin Press, 2019, pp. 203-209.

[50] Hu Jianmiao Expert Studio. Generative AI Governance Urges Legislative Follow-up [N]. Learning Times, 2024-03-13 (A3).

[51] Xiaodong Xie. Risk and Control: On the Protection of Personal Information in the Application of Generative Artificial Intelligence [J]. Political and Legal Essays, 2023(4): 59-68.

[52] [German] Amine Gruenwald, ed., Handbook of Technology Ethics, translated by Wu Yu, Beijing: Social Sciences Literature Press, 2017, pp. 67-68.

[53] Duan Weiwen. Accurate Assessment of the Social and Ethical Risks of Generative Artificial Intelligence [J]. China Party and Government Cadres Forum, 2023(4): 76-77.

[54] Bai Junyi and Yu Wei: "The Limitations of Educational Artificial Intelligence Ethics Principles and Their Rectification-Including a Discussion on the Principle-Deed-Law Framework", China Electro-Education, No. 2, 2024.

[55] BUBECKS, CHANDRASEKARANV, ELDANR, et al. Sparks of Artificial General Intelligence: Early Experiments with GPT−4 [EB/OL]. (2023−03−22) [2025−02−21]. https://arxiv.org/abs/2303.12712

[56] HANSSON S O. Risk and safety in technology [M]// MEIJERS A. Handbook of the philosophy of science, philosophy of technology and engineering sciences, Oxford: Elsevier, 2009: 1069-1102.

[57] Zhao Zhiyun, Xu Feng, Gao Fang, et al. Several understandings on the ethical risks of artificial intelligence [J]. China Soft Science, 2021(6): 1-12.

[58] Frangois Candelon, Lisa Krayer, Saran Rajendran, et al., "How People Can Create — and Destroy — Value with Generative AI", BCG Henderson Institute Website https://www.bcg.com/publications/2023/how-people-create-anddestroy-value-with-gen-ai (Oct. 3, 2024).

[59] Chen Yongwei. "Beyond ChatGPT: Opportunities, Risks and Challenges of Generative AI" [J]. Journal of Shandong University (Philosophy and Social Sciences Edition), 2023(3): 127-143.

[60] Wang Xiaodong: Critique of Intersubjectivity Theory in Western Philosophy, Beijing: China Social Sciences Press, 2004, p. 22.

[61] Wu Qiantao, Xu Baicai et al., Theory and Practice of Ideological and Political Education in Colleges and Universities, Beijing: Peoples Publishing House, 2012, p. 30.

[62] FLORIDI L, SANDERS J W. On the Morality of Artificial Agents [J]. Minds and Machines, 2004(3): 349−379.

[63] Jia Luman, Chen Fan. The reconstruction of moral subject under technical mediation [J]. Research on Dialectics of Nature, 2018(6): 39-43.

[64] COECKELLBERGH M. David J. Gunkel: The Machine Question: Critical Perspectives on AI, Robots, and Ethics [J]. Ethics and Information Technology, 2013(3): 235−238.

[65] Zhang Lu. Risk Governance and Supervision of General Artificial Intelligence: Issues and Challenges Triggered by ChatGPT. E-Government, 2023(9): 14-24.

[66] Ye Ying, Zhu Xiuzhu et al., From the Outbreak of ChatGPT to the Enlightenment of GPT Technology Revolution, Intelligence Theory and Practice, No. 6, 2023.

[67] Wang Shaoshao. The Obstacles of Generative Artificial Intelligence in Academic Misconduct Governance and Countermeasures [J]. Science Studies, 2024(7): 1361-1368.

[68] Xu Lan, Wei Qingyi, and Yan Yi. Strategies and Principles for the Use of Generative Artificial Intelligence in Higher Education from an Academic Ethics Perspective [J]. Educational Development Research, 2023(19): 49-60.

[69] Chen Changfeng and Huang Yangkun: The Knowledge Function of ChatGPT and the Crisis of Human Knowledge, Modern Publishing, No. 6, 2023.

[70] Jonathan J. Tennant, "Web of Science and Scopus Are Not Global Databases of Knowledge", European Science Editing, no. 46, 2020.

[71] "Tracking of Student Information Leakage Incident in Study Connect; Some Sellers Sold Overnight, Claiming to Have Been Bought Out by Major Investors", China News Network: https://www.chinanews.com.cn/cj/2022/06-23/9786718.shtml (September 25, 2024).

[72] Qu Haowei and Li Junshu: On the Legal Protection of Data Security for Virtual Digital Human Applications--Taking the Implementation of Generative Artificial Intelligence as the Approach, Electronic Intellectual Property, 2024, No. 7.

[73] Xiao Dong Xie. Data Security Risks and Adaptive Governance in Generative Artificial Intelligence [J]. Dongfang Methodology, 2023(5): 106-116.

[74] FARINAM, YUX, LAVAZZAA. Ethical considerations and policy interventions concerning the impact of generative AI tools in th e econom y an d i n societ y [J] AI and ethics, 2025, 5: 737- 745.

[75] Feng Jianjun. Citizen Subjectivity and Its Cultivation [J]. Educational Science, 2020, 36(6): 1-6.

[76] MATTHIAS A. The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata [J]. Ethics and Information Technology, 2004(3): 175−183.

[77] LEBEN D. A Rawlsian Algorithm for Autonomous Vehicles [J]. Ethics and Information Technology, 2017(2): 107−115.

[78] Guo Jing. Responsibility attribution based on human-machine cooperative action [J]. Studies on dialectics of nature, 2020(11): 54-60.

[79] Xie Yu, Wang Xiaoyi. Ethical Risks of Artificial Intelligence Emotion and Its Response [J]. Ethics Research, 2024(1): 132-140.

[80] See Louis et al. V. Safe Rent et al. (D. Mass), Civil Rights Division U.S. Department of Justice. [2024−11−03]. https://www.Justice.gov/crt/case/louis-et-al-v-saferent-et-al-d-mass

[81] Xu Yongchun. United Nations Report: Generative AI Exacerbates Gender Bias [N]. Xinhua Daily Telegraph, 2024-03-09 (012).

[82] Buchanan B, Lohn A, Musser M, et al. Truth, lies, and automation [J]. Center for Security and Emerging technology, 202 1, 1(1): 2.

[83] Kreps S, Kriner D. How AI threatens democracy [J]. Journal of Democracy, 2023, 34(4): 122−131.

[84] Zhao Juecheng, a heavy-duty task ahead! U.S. military moves: Accelerating tech partnerships with American tech giants to militarize generative AI [EB/OL]. Global Times, January 30, 2024 [November 16, 2024]. https://world.huanqiu.com/article/4GNTglNUyew?re=nextnews

[85] OPC launches investigation into ChatGPT, OPC. (2023−04−04) [2024−11−06]. https://www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230404/

[86] Philip N. Yannella, Timothy W. Dickens. OpenAI Class Action Likely to Increase Scrutiny of Webscraping and Data Collection Practices [EB/OL]. (2023−07−06) [2024−11−09]. https://www.ballardspahr.com/insights/alerts-and-articles/2023/07/openai-class-action-likely-to-increase-scrutiny-on-webscraping

[87] Wang Qinghua, Hu Lintian. The Technical and Legal Construction of the Responsibility Mechanism for Generative Artificial Intelligence [J]. China Law Review, 2024(4): 114-130.

[88] Wach K, Duong C D, Ejdys J, et al. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT [J]. Entrepreneurial Business and Economics Review, 2023, 11(2): 7−30.

[89] Frances D'Emilio, Matt O'Brien. Italy temporarily blocks ChatGPT over privacy concerns [EB/OL]. (2023−03−31) [2024−11−10]. https://www.Pbs.org/newshour/science/italy-temporarily-blocks-chatgpt-over-privacy-concerns

[90] ChatGPT fined $ 2,829 for exposing personal info of 687 Korean users [EB/OL]. (2023 −07−27) [2024−11−10]. https://www.koreatimes.co.kr / business/companies /20230727 /chatgpt-fined-2829-for-exposing-personal-info-of-687-korean-users

[91] Li Yan. High Attention to the "Intelligence Gap" in the AI Era [EB/OL]. Global Times, (2023-04-21) [2024-11-12]. https://opinion.huanqiu.com/ article /4CZgHgDqRno

[92] See Eric Revell. Microsoft tells European regulators Google has an edge in generative AI [EB / OL]. (2024−03−17) [2024−11−13]. https://www.Foxbusiness.com/technology/microsoft-tells-european-regulators-google-has-edge-generative-ai

[93] Staff in the Bureau of Competition & Office of Technology. Generative AI Raises Competition Concerns [EB/OL]. (2023−06−29) [2024−11−16]. https://www.Ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns

[94] [German] Ulrich Beck: Risk Society--A New Road to Modernization, translated by Zhang Wenjie et al., Nanjing: Yilin Press, 2020, p. 24.

[95] Wu Guosheng. Classics of Philosophy of Technology [M]. Beijing: Tsinghua University Press, 2022: 6.

[96] Jia Luman and Chen Fan: "Research on the Reconstruction of Moral Subjects under Technical Mediation", Dialectics of Nature Research, No. 6, 2018.

[97] Li Mang and Yu Luyao: "The Enlightenment of Learning from the Past to the Digitization of Education", Electric Education Research, No. 4, 2024.

[98] Zhang Hongzheng: The Inevitability and Controllability of Technological Alienation from the Perspective of the Duality of Technology, Science, Technology and Dialectics, No. 5, 2005.

[99] Qiao Ruijin, Mou Huan Sen et al., Introduction to Philosophy of Technology, Beijing: Higher Education Press, 2009, p. 153.

[100] Liu Jinrui. New Risks and Regulatory Framework for Generative AI Large Models [J]. Research on Administrative Law, 2024(2): 17-32.

[101] Miao Fengchun, Holmes Wayne, Huang Ronghuai, et al., "AI and Education: Guidance for Policy-Makers", https://unesdoc.unesco.org/ark:/48223/pf0000376709 (Sep. 23, 2024).

[102] See Japans Ministry of Education, Culture, Sports, Science and Technology. "Utilization of Generative Artificial Intelligence in Elementary and Secondary Education (Ver.2.0) (Published on December 26, Reiwa 6 [EB/OL]. [2024-11-16]. https://www.mext.org/Go.jp/a_menu/other/mext_02412.html

[103] Chen Weixing, The Epistemological Challenge of Intelligent Communication, International Journal of Journalism, No. 9, 2021.

[104] [Alexander of Greece] Nicomachean Ethics, translated by Deng Anqing, Beijing: Peoples Publishing House, 2010, p. 38.

[105] Chen Fan and Li Jiawei. Technology as the Other: New Considerations on the Ethical Relationship between Humans and Technology [J]. Journal of Wuhan University (Philosophy and Social Sciences Edition), 2022, 75(6): 50-59.

[106] Cheng Le. Regulatory Appraisal of General Artificial Intelligence under the Perspective of "Digital Humanism" [J]. Political and Legal Essays, 2024(3): 3-20.

[107] Chen Yingda, Wang Wei. From "Urgent Use First" to "Gradual Improvement": The Construction of Generative Artificial Intelligence Governance System [J]. E-government, 2024(4): 113-124.

[108] Cheng Le. The Situation, Challenges and Prospects of Generative Artificial Intelligence Governance [J]. Peoples Forum, 2024(2): 76-81.

[109] Zhang Linghan. What Kind of AI Law Does China Need? — The Basic Logic and Institutional Framework of Chinas AI Legislation [J]. Legal Science (Journal of Northwest University of Political Science and Law), 2024, 42(3): 3-17.

[110] Carlos Moreira and David Ferguson. The Code of the Superhuman [M]. Zhang Yi (Translated). Beijing: CITIC Publishing Group, 2021: 4.

[111] Carlos Moreira and David Ferguson. The Code of the Superhuman [M]. Zhang Yi, translator. Beijing: CITIC Publishing Group, 2021: 10.

[112] The Ministry of Industry and Information Technology of the Peoples Republic of China. Interim Measures for the Administration of Generative Artificial Intelligence Services [EB/OL] (2023-08-30) (2024-12-02). https://<1>.cn/gyhxxhb/jgsj/cyzcyfgs/bmgz/xxtxl/art/2023/art_4248f433b62143d8a0222a7db8873822.html

[113] KENTHAPADIK, LAKKARAJUH, RAJANIN. Generative AI meets responsible AI: practical challenges and opportunities [C] // Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023: 5805-5806.

[114] Notice on the Issuance of the Measures for the Review of Scientific and Technological Ethics (Trial) [EB/OL] (2023-09-07) (2024-12-02). https://www.gov.cn/gongbao/2023/issue_10826/202311/content_6915814.html

[115] Cao Jianfeng. The Pathways and Implications of AI Ethics and Governance in the EU [J]. Artificial Intelligence, 2019(4): 39-47.

[116] Xue Lan. How to Build a Balanced and Inclusive AI Governance System [EB/OL] (2021-12-11) [2024-03-17]. https://mp.weixin.qq.com/s/vAdOea-IX-3aBQOosKPXiQ]

[117] Global Forum on Artificial Intelligence Governance Framework [EB/OL]. (February 7, 2024) [December 20, 2024]. https://tech.huanqiu.com/article/4GUG3d5zcFm