ISSN: 2710-0170

A New Paradigm for Ethically-Oriented Agile Governance of Generative Artificial Intelligence

Yujian Zhao

Beijing Normal University, Beijing, 100875, China

Abstract

Ethically-oriented agile governance for generative artificial intelligence represents the optimal governance model for responding swiftly and dynamically to ethical risks associated with GenAI. It provides theoretical underpinnings for establishing a systematic ethical governance framework across the entire GenAI ecosystem. Presently, this governance model is confronted with an inevitable transition from theoretical exploration to practical implementation pathways. Accordingly, the practical implementation of ethically-oriented agile governance should take into account three core dimensions: first, establishing ethical consensus principles and scenario-specific standardized guidelines as governance instruments; second, embedding data ethics, algorithmic ethics, and review-evaluation mechanisms throughout the entire lifecycle of intelligent systems during the governance intervention phase; and third, facilitating constructive interaction among governance entities at the macro, meso, and micro levels. In line with the principles of scientific and technological innovation, a robust science and technology ethics framework that is compatible with China's national context should be established.

Keywords

Generative Artificial Intelligence, Agile Governance, Self-Regulatory Awareness, Technology Ethics Framework.

1. Introduction

The year 2025 marks a pivotal turning point for China's Generative Artificial Intelligence(GenAI)technology , which is witnessing explosive growth. DeepSeek has emerged as a "technical maverick" in China's vertical domains by overcoming the trillion-parameter barrier at one-tenth the training cost. Meanwhile, through its parallel processing technology spanning the entire workflow "data crawling → code generation → system deployment", Manus has become the "commander" for complex task chains. Presently, GenAI has been deeply integrated with myriad industries via its versatility, multimodality, and emergent capabilities, fundamentally reshaping the relational structure of the "human-technology-world" triad.As post-phenomenological philosopher Peter-Paul Verbeek argues, with the evolution of human-machine relationship models[1]—from the cyborg relationship to the composite relationship—the dynamic between humans and machines has shifted from a "master-servant" paradigm where machines assist humans to a "symbiotic" paradigm where machines supplant human decision-making. In this era of intelligent societal transformation, ethical transgressions triggered by GenAI have become a global governance challenge, sparking intense debates among experts and scholars worldwide regarding the value norms and ethical dilemmas associated with GenAI.In this era of intelligent societal transformation, ethical transgressions triggered by GenAI have become a global governance challenge, sparking intense debates among experts and scholars worldwide regarding the value norms and ethical dilemmas associated with GenAI.

ISSN: 2710-0170

In an era of rapid technological iteration and evolving ethical risks, we must objectively recognise the flexibility, dynamism, long-term nature, and complexity of moral and ethical hazards.[2]We should establish a dynamic adjustment governance mechanism based on agile coordination to balance AI development and safety, thereby fostering a fair, transparent, and sustainable ethical ecosystem for artificial intelligence. This paper examines the ethical challenges and underlying dynamics of generative artificial intelligence, summarising the current governance dilemmas in this field, and further explores a new paradigm of agile ethical governance that is more suitable and effective for the practical operations of GenAI—including its design, development, and application—to adapt to the emerging trends and requirements of global artificial intelligence technology development in the future.

2. Ethical Challenges and Underlying Mechanisms of GenAI

2.1. Risk of Ethical Value Imbalance: Mismatch Between Technological Development and Human Core Values

2.1.1. Indiscriminate Learning Techniques Amplify Inherent Value Biases Within Human Society

The vast quantities of data input during the training and application phases of GenAI are automatically and indiscriminately scraped and collected by machines. This low-quality, undiverse data carries deep-rooted biases and discrimination inherent in human society. Such biased and discriminatory data predominantly centres on sensitive identity markers—including gender, ethnicity, employment status, age, geography, consumption patterns, and discrimination against vulnerable groups—embedding these prejudices throughout the entire lifecycle of GenAI: from research and development to deployment and application. This inherent structure renders GenAI incapable of escaping the infiltration of human discrimination and bias. Societal stereotypes, prejudices, and discriminatory viewpoints find indiscriminate inheritance and expression through GenAI as a new medium. This accelerates the transmission speed and dissemination scope of discriminatory and biased data, amplifying societal 'flaws' and running counter to core ethical values such as 'impartiality, non-discrimination, fairness, justice, and neutrality.' Consequently, it transforms into dangerous 'black technology.'[3]

2.1.2. Algorithm Design and Application Reinforce Inherent Value Biases in Human Society

The various risks arising from algorithmic applications all point to the underlying power structures of algorithms. The risks of value imbalance induced by algorithms primarily manifest in three aspects: Firstly, during algorithm design, designers may unconsciously create unfair distinctions based on identity characteristics such as race and gender.[4]During training, when algorithms automate decision-making by supplanting human judgement, they may generate discriminatory algorithmic decisions or content outputs; [5]Secondly, illicit actors may exploit 'algorithm-driven addictive recommendations' or 'discriminatory pricing techniques' by integrating users' value preferences, lifestyle habits, and social patterns to deliver targeted information, thereby restricting the convergence of diverse viewpoints and ideas.[6]This leads to new societal risks in the intelligent era, such as narrowed human cognition, big data 'loyal customer price discrimination', and 'information cocoons'. Thirdly, the algorithmic black box obscures the operational logic and decision-making basis of large models, reproducing and extending inherent biases and discrimination within human society.[7]Moreover, discriminatory outcomes from algorithmic decisions can propagate through subsequent decisions, creating a chain reaction that deepens societal 'inequality' and 'opacity' while challenging fundamental human dignity.

ISSN: 2710-0170

2.2. Risk of Ethical Norms Running Amok: The Contradiction Between Technological Innovation's Proactive Nature and Ethical Norms' Lagging Nature

2.2.1. The Collapse of Human Privacy Space

GenAI can collect, record, share, and analyse vast amounts of personal information, log data, and symbolic information, severely compressing the space where privacy exists. Furthermore, GenAI may infer that data collected during human-machine interactions constitutes private or sensitive information, inadvertently generating outputs containing genuine personal details during creative processes, thereby infringing upon individual privacy rights.[8]Such occurrences happen constantly whenever users open and utilise apps, whether uploading personal information, corporate trade secrets, critical code, habitually tagging locations with personal details, or inadvertently clicking buttons permitting intelligent systems to access and gather personal data.[9]This gradually transforms the 'society of strangers' with its clear public-private boundaries into a 'society of transparent individuals' where everything is laid bare. It severely undermines the fundamental value of privacy as a core aspect of social identity and poses unprecedented challenges to legal protections for privacy rights.

2.2.2. Triggering a Dispute over Subjecthood Between Humans and Intelligent Agents

GenAI possesses increasingly high levels of autonomy, capable of making independent decisions and taking actions without explicit human instructions, demonstrating understanding and autonomous decision-making abilities comparable to those of humans. Automated decision-making technologies subvert the ethical paradigm of the 'subject-object duality' relationship,[10]shifting from a human-centric model where 'humans are within the decision-making circle' to one where 'humans are outside the decision-making circle' as AI agents assume autonomous decision-making.[11]This shift severely disrupts the ecological fabric of human social interaction.[12]This has fostered a human-machine dominance relationship where GenAI either usurps the master's role or blurs the distinction between master and servant. If Kant's assertion that 'man legislates for nature' signified humanity's emancipation from natural constraints and the establishment of human subjectivity, then the advancement of artificial intelligence now threatens this very status of subjectivity. [13]Today, the question of whether to grant GenAI subject status has become a central and long-standing point of contention in both theoretical and practical circles, triggering confusion over identity recognition. The primary reasons are twofold: on the one hand, future intelligent entities will inevitably possess heightened self-awareness and independent reasoning capabilities. We must anticipate the various societal issues arising when the 'singularity' arrives and formulate effective solutions in advance. On the other hand, the legal subject for assuming tort liability remains undetermined, and tort compensation liability cannot yet be reasonably allocated.

2.3. Risk of Ethical Imbalance: Impact on Human Agency and Social Order

2.3.1. Diminished Human Agency

The blurring of human-machine boundaries and intensifying human-machine conflicts have gradually diverted human-centred ethical principles from their original course within GenAI applications. On one hand, technology firms employ algorithmic recommendation systems to meticulously analyse and deeply mine user preferences, embedding these insights into algorithms. They then refine recommendation schemes through user experience feedback and usage patterns, pursuing commercial profit as their developmental objective. Within the fortress of algorithmic satisfaction, users progressively lose exposure to experiences beyond their cognitive horizons. The 'emotional intervention' of algorithmic recommendation technologies deprives humans of opportunities to explore novel experiences, significantly

ISSN: 2710-0170

diminishing autonomous agency. [14] On the other hand, human self-awareness faces disruption. In certain domains, users can leverage GenAI to accomplish specialised tasks and refine outputs through the technology's self-feedback mechanisms.[15]This leads users to neglect their own capacity for innovation, practical application, and training, gradually eroding their ability for independent thought, critical reasoning, and resource integration—ultimately impacting societal development.[16]Particularly within academia, researchers employing ChatGPT to generate scholarly papers for publication in authoritative journals violate academic integrity and cast doubt upon the academic standards of these journals.

2.3.2. Human Overdependence on Intelligent Agents

As GenAI advances in both interactive capabilities and professional expertise, humans will grow accustomed to intelligent systems participating in their work and daily lives. They may cease to question the authenticity of content generated by large models, and may even develop emotional dependencies due to prolonged provision of emotional value by these agents.[17]This may adversely affect interpersonal relationships and mental wellbeing. Furthermore, emotional dependency manifests in GenAI technologies that revive the deceased by highly replicating their appearance, behaviour, and voice to provide users with emotional solace. This virtual 'overlay of life and death' leads users to mistakenly believe the deceased are still alive, posing significant challenges to traditional human culture, religious beliefs, and perspectives on life and death.

2.3.3. The Alienation of Human Labour Relations

Marx observed: '… when men begin to produce the means of life necessary for themselves … they begin to distinguish themselves from animals.'[18] Labour, as an essential attribute of humanity, constitutes a defining characteristic distinguishing humans from animals. To enhance their existence, humans create wealth, comprehend the world, and transform themselves through labour.[19]Artificial intelligence represents an extension of human intelligence, conceived precisely to serve and benefit humanity. However, GenAI has propelled artificial intelligence beyond specialised domains into general-purpose applications, endowing it with the capacity to simultaneously address or accomplish multiple tasks. This advancement has led to the displacement of low-end, repetitive labour positions and language-based task-oriented industries by AI systems capable of executing identical tasks with high efficiency and low cost.[20]Furthermore, GenAI is widely utilised by teachers and students in education. Some pupils employ intelligent software to complete assignments, stifling their capacity for independent innovation and critical thinking. Frequent interaction between adolescents and AI may also pose mental health challenges for young people.[21]

2.3.4. Creating a Significant Number of Hidden "Digital Refugees"

Differences in technological learning abilities among various groups may lead to disparities in their capacity to utilise technology. The societal digitalisation process has given rise to the phenomenon of the 'digital divide'. The emergent and iterative nature of generative artificial intelligence has further deepened and widened this divide, exacerbating inequalities in developmental rights and living conditions across humanity in areas such as social life, economic activity, education and culture, and welfare provision.[22]The covert nature of the digital-intellectual divide renders its manifestations difficult for individuals to perceive during algorithmic operations, inadvertently creating vast numbers of hidden 'digital refugees'. Differences in individuals' data ownership, application capabilities, and innovation capacity readily create information gaps and wealth disparities between privileged and disadvantaged groups, leaving the latter vulnerable to intellectual obsolescence. This phenomenon not only impedes individuals' access to information and participation in social activities but also reshapes social structures and individual status to some extent. It further touches upon profound ethical and philosophical questions, involving ambiguities at the margins of ethics. It

ISSN: 2710-0170

may also undermine the universal humanist stance, significantly challenging fundamental values upon which humanity relies—such as the labour value, survival value, and spiritual sustenance that underpin existence.[23]

3. Ethical Governance Dilemmas in Generative Artificial Intelligence

3.1. Immature Technological Development Pathways and Operational Mechanisms

From the operational mechanisms and technological pathways of GenAI, the development of large language models presents significant uncertainty. On the one hand, from OpenAI's GPT-4.0 to DeepSeek's DeepSeek, and even the forthcoming GPT-5, the deep learning capabilities of large language models have demonstrated exponential growth, far exceeding human projections. They now exhibit comprehension and logical reasoning abilities comparable to human intelligence. However, the technological trajectory remains unclear, facing numerous technical bottlenecks such as slow data updates, insufficient data scraping, susceptibility to 'hallucinations', limited generative capabilities, algorithmic black boxes, and inadequate computational power. GenAI technology has yet to attain cognitive and reasoning capacities surpassing human rationality, and the timing of the 'singularity' remains unknown. Consequently, the uncertainty surrounding future technological directions prevents humans from effectively predicting and responding to the ethical risks posed by intelligent agents. Conversely, AI technology exhibits a characteristic where more advanced operational mechanisms correlate with slower rates of intelligence enhancement. Take autonomous driving as an example: current technology remains at Level 4, [24]still some distance from Level 5,[25] requiring further research and development to bridge this gap.[26]

3.2. Embedding Ethics into Algorithms Cannot Mitigate All Ethical Risks

The regulatory approach of embedding ethics into algorithms involves utilising coding techniques to incorporate moral algorithms (resembling human ethical consciousness) within artificial intelligence systems or products. These moral algorithms guide AI to make actions and choices aligned with human ethical standards.[27]While theoretically viable, this regulatory approach presents practical shortcomings. On the one hand, artificial intelligence is embedded within a triadic 'human-society-nature' interaction system[28] and is influenced by diverse factors such as politics, economics, culture, environment, customs, and cognition across different eras and regions. Consequently, when collecting data reflecting specific worldviews and values, the inherent value biases or certain ethical concepts within human society may become entrenched and amplified due to an excessive pursuit of the instrumental rationality of technological tools. On the other hand, dynamic intelligent systems operating autonomously and creatively possess characteristics of ubiquity and hallucinatory generation. In this respect, artificial intelligence shares similarities with human intelligence. However, human ethics and morality are 'constructed by rational agents through dialectical reflection on their own agency'.[29] Actors possess rational self-reflection and self-examination capabilities, consciously and proactively practising ethical requirements in pursuit of 'truth, goodness, and beauty'. In contrast, as Coeckelbergh observes, artificial intelligence appears devoid of emotions such as love, sympathy, or concern[30],and lacks the capacity to autonomously select ethically compliant behaviour driven by factors like dignity, achievement, or honour.[31]

3.3. Technological Illusions and Human Cognitive Limitations Undermine Existing Ethical Norms

Human modes of thought and cognitive patterns are inherently limited. Human judgement is constrained by the completeness, accuracy, and timeliness of information, and is also influenced by specific cultural contexts, social contexts, and value systems.[32]This inherently

ISSN: 2710-0170

limits humans' capacity to recognise the manifestations, severity, and probability of technological risks, thereby hindering their ability to identify genuine hazards. Assessments of ethical risks by different individuals may also suffer from evaluation biases, errors, or divergent viewpoints. In the current era of exponential iteration in intelligent technologies, traditional ethical risks may no longer be pertinent, while contemporary risks might remain unrecognised or unperceived by humanity. Consequently, human limitations in risk perception may lead to inaccuracies in formulating ethical principles or their inappropriate application. At present, the specific content and interpretations of ethical principles proposed by researchers lack consistency. No consensus has been reached on foundational ethical principles for GenAI, nor have existing principles been sufficiently elaborated, creating potential interpretative conflicts. Even if consensus on foundational principles were achieved, no single ethical principle can universally address all ethical scenarios.[33]

4. Systematic Development of an Ethical Governance Framework for the GenAI Ecosystem

4.1. Establishing Ethical Consensus Principles and Scenario-Specific Standardisation Guidelines

4.1.1. Formulating Universal Principles Applicable Across Development, Training, and Deployment Stages

The outcomes of the digital technology and artificial intelligence revolution remain unpredictable, for history cannot reveal what constitutes the ultimate technological revolution. From Asimov's 'Three Laws of Robotics' proposed in 1940 to the formal introduction of 'Robotics Ethics' at the inaugural International Symposium on Robotics Ethics in 2004[34],the direct threats, structural imbalances, and tensions arising from artificial intelligence have driven an explosive adjustment of ethical principles worldwide. Ethical principles such as human-centredness and fairness are being deepened within international cooperation mechanisms, forming consensus-based principles for AI ethics.[35]International Organisations[36],National and Regional Authorities[37]have successively introduced ethical regulatory policies aligned with their AI governance philosophies and technological development stages to balance ethical governance with technological innovation. These include market- and innovation-oriented regulatory models exemplified by the United States[38],as well as proactive intervention models represented by the European Union[39],and neutral models balancing development and security championed by China.[40]Ethical governance of GenAI should encompass the entire lifecycle from R&D to application, ensuring that data collection and organisation, model training and validation, application evaluation and feedback[41],and all other stages are guided by socially agreed technological ethics principles and ethical norms.[42]

4.1.2. Establishing Standardised Guidelines for Diverse Application Scenarios

Within the philosophical domain, British philosopher Jonathan Dancy advocates moral particularism(a philosophical theory). He contends that each ethical issue is unique, requiring individual consideration, and that specific moral principles cannot fully determine human moral thinking and judgment.[43]Specifically, the same rationale may prompt different actions across varying contexts, while identical actions may stem from distinct motivations in different settings. A particular consideration may be supported in one scenario yet opposed in another, or even deemed wholly unreasonable. From this philosophical-ethical perspective, moral particularism aligns with artificial intelligence ethics. Artificial intelligence products are extensively deployed across industries. The same language model may foster emotional dependency in conversational settings yet risk privacy violations in workplace contexts. Even

ISSN: 2710-0170

within identical scenarios, the same chatbot may raise distinct ethical concerns for varying reasons. Consequently, drawing on the theory of moral particularism provides a framework for addressing the contextual ethical dilemmas of AI, enabling the development of scenario-specific guidelines grounded in consensus principles.

Standardisation of application scenarios serves as a vital driver for industry development and a key element in AI ethical governance. China has established a standardisation system grounded in standardisation organisations, the content and type distribution of standards, and standard-drafting bodies. However, the development of standards for AI ethical governance has yet to keep pace with industrial and technological advancements. Specific regulations—including product standards, technical standards,and management standards—must be formulated for high-risk domains and scenarios such as education, healthcare, and autonomous driving, thereby deepening the regulatory role of standards in technology.

It should be noted that as GenAI maintains rapid technological advancement, the associated ethical risks across diverse application scenarios evolve continuously, necessitating ongoing refinement of ethical norms. This demands a thorough understanding of AI's potential and limitations, alongside the continuous optimisation and dynamic adjustment of ethical standards.[44]Therefore, diverse societal actors should be permitted to identify novel ethical standards throughout the entire process from R&D design to application. They should be able to articulate differing attitudes and interpretative approaches towards technological ethics, thereby facilitating the formation of consensus-based ethical principles.[45]These insights should then be fed back into the public repository of ethical standards, thereby systematically advancing the healthy and sustainable development of artificial intelligence.

4.2. Embedding Data Ethics, Algorithmic Ethics, and Review Mechanisms Throughout the Entire Lifecycle of Intelligent Systems

4.2.1. Integrating Data Ethics Governance Across the Entire Process of R&D, Training, and Application Phases

Data constitutes the ‘raw material’ for GenAI, primarily comprising training data and user data, manifested in diverse formats manifested in diverse formats such as text, images, symbols, audio, and video. Large language models learn human ethical issues embedded within data, and as application scenarios expand, these ethical concerns are further extended and propagated.Consequently, data ethics governance must prioritise the lawful acquisition of data, the ethical compliance of supplied content, and the review and assessment during training iterations and deployment.

During the initial R&D phase, collect traceable, legitimate, and standardised public data, as well as unprocessed raw data from specialised databases, to establish and regularly update training corpora. This reduces the model's exposure to content inconsistent with core socialist values at the data acquisition source, thereby reducing its potential impact on societal ethical frameworks. In addition, professional data management agencies may be entrusted to conduct segmented supervision throughout the entire process of data collection, storage, management, and utilisation.

During the design and development phase: Firstly, employ dynamic monitoring and evaluation mechanisms to scrutinise data quality for misleading or discriminatory ethical issues. Establish negative lists for anomalous data, along with corresponding processing protocols and solutions, to ensure that training data aligns with ethical frameworks and moral standards. Secondly, embed core values such as ssafeguarding human well-being and building trustworthy AI within foundational technologies. This enhances data integrity, timeliness, consistency, standardisation, and accuracy.[46]Thirdly, manually annotate privacy-sensitive data and information involving patents, copyrights, and other similar matters[47]to prevent copyright infringement and data security risks.[48]

ISSN: 2710-0170

During deployment and application phases, configure user settings to avoid privacy violations and prevent data leaks, and ensure that application systems do not crawl sensitive information proactively provided by users.[49]When utilising information involving public discourse and non-authoritative expressions, careful identification and strict screening of sensitive data must be carried out to prevent the secondary dissemination of such information from amplifying adverse impacts after large-model training .

4.2.2. Embedding Algorithmic Ethics Governance Throughout the Entire Process of R&D, Training, and Application Stages

Algorithms serve as the 'engine' of GenAI, determining how model systems process and utilise data to make more intelligent decisions. The black-box nature of algorithms necessitates that algorithmic governance prioritise transparency, interpretability, and reliability.

During the design, development, and training phases, ethical norms and value principles should be embedded in algorithmic models, and mechanisms for correcting value deviations should be established. Within the model, eethical review, testing, and evaluation mechanisms—where algorithms supervise algorithms—should be implemented: clear review criteria should be defined, and virtual testing should be conducted through algorithmic simulation environments to identify potential biases and discrimination in algorithmic design, enabling full traceability and impartial review of algorithmic logic. By establishing incentive mechanisms, encourage machines to verify compliance with ethical principles in each learning iteration, and provide positive feedback for adjustments to algorithms and training modes to guide outputs that align with human moral expectations .[50]Additionally, developers may establish registration and filing systems to facilitate traceable reviews of critical information such as large model parameters, algorithms, and data. [51]

During deployment and application, priority should be given to expanding channels for non-technical personnel and the general public to comprehend algorithmic logic. This may involve developing algorithmic auditing tools specifically designed for non-technical users, as well as recommending professional crowdsourced auditing services or proxy auditing firms to the public. Such measures would facilitate ordinary citizens' assessment of the rationality and potential societal impacts of generative AI decision-making.[52]Additionally, government authorities should establish open user feedback and review platforms, which should connect with relevant industry organisations and technology companies to enable users who have experienced unfair treatment by intelligent systems to directly communicate with technology firms through appeals. This approach not only provides users with avenues for remedy but also compels algorithm designers to engage in continuous self-reflection and optimisation.

4.2.3. Embedding Ethical Risk Review and Assessment Mechanisms Throughout the Entire Process of R&D, Training, and Application Stages

The primary function of scientific and technological ethics review mechanisms lies in detecting and identifying the presence of serious ethical risks in scientific and technological activities. Their scope of application primarily encompasses specific innovative scientific and technological endeavours at the practical level.[53]Academic communities and technology enterprises have begun implementing specific measures, including ethical self-review by AI researchers, peer review, and institutional review. However, these measures still require further refinement and improvement.[54]

Prior to the market release of GenAI products, ethical considerations may be integrated into product quality evaluation systems to conduct ethical bias testing.[55] If improper use by users causes the product's outputs to violate social ethical norms, administrative penalties such as warnings, usage restrictions, or fines may be imposed on users to deter and caution them regarding the proper use of intelligent applications.

ISSN: 2710-0170

When risks occur, pre-established contingency response plans shall be activated immediately. A technology ethics risk monitoring and early warning body shall be established to incorporate professionals from institutions of higher education, research organisations, social groups, and technology enterprises. This body shall conduct comprehensive assessments of potential rule conflicts, societal risks, and ethical challenges across diverse scenarios and propose countermeasures .[56]

Establish incentive mechanisms that prioritise ethical review. For similar projects, if researchers have conducted detailed ethical risk analyses and implemented preventive measures, and have achieved better results than other similar projects during development and application, subsequent research endeavours should receive prioritised and rolling funding. During ethical review, proposals should not be summarily rejected simply because researchers have identified many potential major ethical concerns in their applications; doing so would fundamentally undermine the purpose of ethical scrutiny.[57]

4.3. Coordinated Interaction Among Governance Entities at Macro, Meso, and Micro Levels

4.3.1. Macro Level: Exercising Government Regulatory Functions

As the principal formulator of regulatory policies, the government should innovate regulatory tools and approaches, systematically establishing and refining ethical norms and standards for science and technology. Firstly, clarify the ethical review and regulatory responsibilities of the Science and Technology Ethics Review Committee operating under a central-local dual management system. According to their respective duties, they shall be responsible for formulating ethical norms, conducting reviews and oversight, and undertaking publicity and education initiatives. Strengthen supervision in sensitive areas such as cultural data security and personal privacy protection, rigorously cracking down on illegal activities including data misuse, privacy breaches, and deepfakes to safeguard citizens' cultural rights and digital intelligence information security.[58]Establish an expert review mechanism for ethical review outcomes of high-risk scientific and technological activities, organise investigations and handling of major scientific and technological ethics cases, and utilise typical cases to enhance warning education.[59]Secondly, keeping pace with the cutting edge of GenAI technology, government-affiliated research institutions should intensify research on the ethics of large models, proactively lay the groundwork for ethical regulatory systems, and introduce the concepts of'sector-specific regulation'and'scenario-based regulation'. Advanced technologies should empower governmental oversight methods to enhance the scientific and intelligent dimensions of governance, while improving the flexibility and applicability of ethical regulatory frameworks. Thirdly, traditional hierarchical governmental organisational structures should be restructured to foster a more streamlined governance architecture.[60]This will foster a cross-departmental, full-cycle, and systematic ethical regulatory framework.[61]Finally, incorporate technical specialists—encompassing operational technicians and industry experts—to foster regulatory tool innovations such as regulatory sandboxes and the 'A.I.Verify' toolkit.

4.3.2. Meso-level: Collaborative Interaction Between Enterprises and Industry Organisations

Strengthening Self-Regulation of Ethical Conduct Among Technology Enterprises

Since 2018, major domestic and international technology enterprises such as Google, Microsoft, IBM, Tencent, and Baidu have taken the lead in ethical governance. They have successively introduced AI governance principles, guidelines, and reports tailored to their respective management frameworks, while establishing dedicated AI ethics governance bodies to actively fulfil their responsibilities in AI governance. [62]As pivotal platforms for developing, deploying, and servicing GenAI, technology enterprises must bolster ethical governance through

ISSN: 2710-0170

heightened ethical awareness, robust mechanism design, and systematic implementation. Firstly, embedding the concept of "ethics first, technology for good" into corporate culture and induction training to subtly guide enterprises and technical personnel towards upholding positive ethical values. Second, while implementing national ethical oversight policies, establish self-regulatory ethical review systems within technology enterprises. Place particular emphasis on institutional development in data security, privacy protection, and user personal information safeguarding to ensure that AI technologies, products, and services comply with trustworthiness requirements. Protect user data security and privacy throughout storage, usage, and destruction processes, preventing incidents such as data misuse or illicit transactions that undermine foundational trust between enterprises and users. Thirdly, to facilitate the effective implementation of ethical self-regulation frameworks, technology enterprises should undertake corresponding organisational restructuring. This includes establishing technical advisory committees and ethics review committees within their organisational structures. These bodies shall review, test, and evaluate whether actions taken at different stages of the large model lifecycle comply with ethical regulatory policies and self-imposed ethical requirements. They shall also propose feasible recommendations and concrete remedial measures to address significant ethical non-compliance issues.

Harnessing the Collaborative Governance Capabilities of Industry Organisations

Achieving precise and effective governance of ethical issues in GenAI cannot rely solely on conceptual policy guidelines and principles. It necessitates the establishment of a comprehensive, systemic regulatory framework involving multi-stakeholder participation and cross-sectoral co-governance. A crucial aspect of future AI development lies in advancing humanity's collective intelligence to address escalating challenges. [63]Therefore, experts from diverse fields should be encouraged to participate in the cultural innovation governance framework for AI, fostering consensus and forming a cohesive governance force. [64，65]]and to propose multi-dimensional, multi-tiered governance strategies to construct a more open and inclusive innovation ecosystem. Relevant industry organisations should actively explore establishing professional associations, fostering cross-sectoral exchanges through conferences and collaborative projects to form a governance framework featuring dynamic oversight both within and beyond the industry. [66]

Regarding intra-industry oversight, industry associations and the media can leverage their professional expertise and organisational resources to conduct effective supervision of intelligent governance practices;[67]industry alliances should undertake the formulation of ethical AI technical standards and management standards, as well as monitoring, certification, and the compilation of case studies. Regarding external oversight, regular consultations are held with external experts from diverse fields including law, philosophy, management studies, and artificial intelligence technology. Enterprises publicly disclose their AI ethical governance practices and research findings to society.[68]Concurrently, through democratised and pluralistic decision-making processes, opinions and suggestions from diverse cultural and social groups are fully solicited and incorporated.[69]This enables dialogue among stakeholders within an environment characterised by comprehensive information, equal opportunities, and transparent procedures, facilitating discussions on relevant AI policy issues to ensure the targeted nature, democratic nature, and transparency of ethical norm-setting .[70]

4.3.3. Micro Level: Enhancing Self-Regulatory Awareness Among Technical Participants

Enhancing Self-Regulatory Awareness Among Developers

Developers possess certain informational and spatial advantages, forming the first line of defence in the ethical governance of GenAI. However, some algorithmic discrimination stems from designers intentionally or unintentionally embedding their own biases into objective

ISSN: 2710-0170

algorithms. Therefore, the ethical literacy and professional capabilities of R&D personnel should be strengthened to achieve refined internal governance.On one hand, enhancing ethical literacy requires comprehensive measures such as public sector ethics awareness campaigns, school education, in-service training, industry seminars, and practical exercises. These should elevate developers' ethical cognition and value judgement capabilities, guiding their ethical decision-making and compliance awareness during development. Concurrently, developers must strengthen self-discipline, proactively conduct self-reviews, enhance self-management, and uphold the fundamental principles of technological ethics. [71]On the other hand, the enhancement of professional capabilities is primarily reflected in overcoming algorithmic black boxes, strengthening algorithmic interpretability, embedding ethical considerations into design, establishing retrospective early-warning mechanisms, and developing self-replicating improvement technologies. It is imperative to cultivate a team of interdisciplinary R&D personnel who possess both technical capabilities and a strong sense of ethical responsibility.

Enhancing Users' Digital Literacy

We should actively implement the requirements for inclusivity and accessibility outlined in the *Ethical Guidelines for New Generation Artificial Intelligence*. Firstly, as the main force driving the nation's future development, the younger generation must be prioritised in cultivating digital literacy, which should be regarded as a key task of the national digital strategy. Cybersecurity education for all citizens and AI ethics education should be incorporated as essential components into the nine-year compulsory education and university curricula. Secondly, users should consciously undergo education in legal norms and ethical principles, adhering to good faith usage, avoiding misuse or abuse, prohibiting unlawful or malicious exploitation, and providing timely, proactive feedback. This will enhance users' information discernment, critical thinking, and moral judgement capabilities. Thirdly, tailored GenAI products and educational programmes should be developed for vulnerable groups such as the elderly and persons with disabilities . Public awareness campaigns on technological ethics should be conducted, encouraging professional societies, associations, and research institutes to establish platforms for disseminating ethical knowledge. Advancing infrastructure development in remote areas and enhancing residents' digital literacy can narrow the "digital divide," bridging knowledge and skill gaps between different groups to ensure that technological innovations equitably benefit all members of society, allowing everyone to share the welfare created by technology. [72]

5. Conclusion

The prevention of ethical risks in artificial intelligence is not a task that can be accomplished overnight, but rather one that requires persistent dedication and sustained long-term endeavour. As technology continues to advance, both the risks inherent within the technology and external risks will continually evolve. Therefore, the ultimate objective of safeguarding against AI ethical risks should be to establish a dynamic and flexible risk prevention mechanism based on human-machine symbiosis. This mechanism would keep risks within the controllable scope of humanity, enabling both AI technology and risk prevention measures to progress and advance in tandem. At present, we are diligently exploring a systematic, resilient, and full-lifecycle risk prevention mechanism that includes specific objectives, approaches, mechanisms, and strategies. Although related research and practice remain in their nascent stages, they lay a foundational groundwork for subsequent research pathways, carrying profound societal significance.[73]Moving forward, ethical governance of GenAI must still address questions such as‘Does code possess moral attributes?’and‘How should ethical codes be formulated in the age of artificial intelligence?’, while further refining the planning, implementation mechanisms, and regulatory measures for ethical governance.

ISSN: 2710-0170

References

[1] Liu Zhiyi, Liang Zheng, Zheng Yejie. Black Mirror and Order: AI Ethics and Governance in a Digital-Intelligent Risk Society [M]. Beijing: TsinghuaUniversity Press, 2022.

[2] Du Yanyong. Ethical Review of Artificial Intelligence: Current Status, Challenges and Prospects [J]. Journal of Donghua University (Social Sciences Edition), 2024, 24(02): 32-39.

[3] Liu Zhiyi, Liang Zheng, Zheng Yejie. Black Mirror and Order: AI Ethics and Governance in a Digital-Intelligent Risk Society [M]. Beijing: Tsinghua University Press, 2022.

[4] Xuefang Xie, Chen Qu. Value Alignment: Research on AI Cultural-Technological Ethical Risks and Pathways for Precise Co-Governance in the AIGC Era [J]. Journal of Lanzhou University (Social Sciences Edition), 2024, 52(03): 147-156.

[5] Future Forum. Artificial Intelligence Ethics and Governance: A Future Perspective [M]. Beijing: People's Posts and Telecommunications Press, 2023: 166.

[6] Luo Yi, Pan Lu. Approaches to Legal Safeguards for Artificial Intelligence Ethics: Conceptual Understanding, Ethical Challenges and Institutional Responses [C]// Rule of Law Culture Anthology 2024, Vol. 1 — Research Collection on Legal Expression of Technological Ethics. School of Environmental Law, Gansu University of Political Science and Law; School of Law, Gansu University of Political Science and Law;, 2024:65-77.

[7] Leprib,OliverN, LtouzeE,Etal.Fair,Transparent,and Accountablle Algorithmic Decision-marking processes, Philsophy & Technology,2018,31:617.

[8] Feng Zixuan. Ethical Stances and Governance Approaches for Generative Artificial Intelligence Applications: The Case of ChatGPT [J]. Journal of East China University of Political Science and Law, 2024, 27(01): 61-71.

[9] Li Yifan, Xu Liang. Examining Ethical Issues in Artificial Intelligence and Corresponding Countermeasures [J]. Foreign Trade and Economic Cooperation, 2024, (05): 27-29+75.

[10] Cai Lin. Logical Justification and Construction Approaches for the Legalisation of Algorithmic Technology Ethics [J]. Legal Science (Journal of Northwest University of Political Science and Law), 2024, 42(04): 92-104.

[11] China Academy of Information and Communications Technology. Research Report on Ethical Governance of Artificial Intelligence (2023) [EB/OL]. (2023-12-26) [2024-12-26]. https://mp.weixin.qq.com/s/EUri4EgJGR9iSTuxMsRmzA.

[12] Tang, Linyao. Legal Regulation of ChatGPT under Embodiment Ethics and the Chinese Pathway [J]. Eastern Law Review, 2023, (03): 34-46.

[13] Liu Zhiyi, Liang Zheng, Zheng Yejie. Black Mirror and Order: AI Ethics and Governance in a Digital-Intelligent Risk Society [M]. Beijing: Tsinghua University Press, 2022.

[14] Li Yifan, Xu Liang. Examining Ethical Issues in Artificial Intelligence and Countermeasures [J]. Foreign Trade and Economic Cooperation, 2024, (05): 27-29+75.

[15] China Academy of Information and Communications Technology. Research Report on Ethical Governance of Artificial Intelligence (2023) [EB/OL]. (2023-12-26) [2024-12-26]. https://mp.weixin.qq.com/s/EUri4EgJGR9iSTuxMsRmzA.

[16] Luo Yi, Pan Lu. Approaches to Legal Safeguards for Artificial Intelligence Ethics: Conceptual Understanding, Ethical Challenges and Institutional Responses [C]// Rule of Law Culture Anthology 2024, Vol. 1 — Research Collection on Legal Expression of Technological Ethics. School of Environmental Law, Gansu University of Political Science and Law; School of Law, Gansu University of Political Science and Law;, 2024:65-77.

[17] China Academy of Information and Communications Technology. Research Report on Ethical Governance of Artificial Intelligence (2023) [EB/OL]. (2023-12-26) [2024-12-26]. https:// mp. weixin. qq.com/s/EUri4EgJGR9iSTuxMsRmzA.

[18] [Germany] Marx, [Germany] Engels. Selected Works of Marx and Engels (Volume 1) [M]. Beijing: People's Publishing House, 2009: 530.

ISSN: 2710-0170

[19] Luo Yi, Pan Lu. Approaches to Legal Safeguards for AI Ethics: Conceptual Understanding, Ethical Challenges and Institutional Responses [C]//Rule of Law Culture Anthology 2024, Vol. 1 – Research Collection on Legal Expression of Technological Ethics. School of Environmental Law, Gansu University of Political Science and Law; School of Law, Gansu University of Political Science and Law;, 2024:65-77.

[20] Zhi, Zhenfeng; Liu, Jiakun. Ethics First: Governance Strategies for Generative Artificial Intelligence [Journal Article]. Yunnan Social Sciences, 2024, (04): 60–71.

[21] China Academy of Information and Communications Technology. Research Report on Ethical Governance of Artificial Intelligence (2023) [Electronic Resource]. (2023-12-26) [2024-12-26]. https://mp.weixin.qq.com/s/EUri4EgJGR9iSTuxMsRmzA.

[22] Future Forum. Artificial Intelligence Ethics and Governance: A Future Perspective [M]. Beijing: People's Posts and Telecommunications Press, 2023: 166.

[23] Feng Zixuan. Ethical Stances and Governance Approaches for Generative AI Applications: The Case of ChatGPT [Journal Article]. Journal of East China University of Political Science and Law, 2024, 27(01): 61-71.

[24] Level 4: Achieving autonomous driving on specific types of roads or under specific weather conditions

[25] Level 5: Capable of autonomous driving in any weather and on any type of road

[26] Khan M A, Sayed H E, Malik S, et al. Level-5 autonomous driving—are we there yet? a review of research literature[J]. ACM Computing Surveys, 2022, 55(2): 1-38.

[27] Dai Yibin. Research Approaches to Artificial Intelligence Ethics [J]. Social Sciences, 2023, (07): 45-52.

[28] Zheng Huangjie. Trustworthy Artificial Intelligence: Governance Foundations for AIGC Amid Technological Ethical Risks [J/OL]. Science and Technology Progress and Policy, 1-11 [2025-09-04]. https://link.cnki.net/urlid/42.1224.G3.20241101.1042.002.

[29] [US] Freeman S. The Collected Papers of John Rawls [M]. Translated by Chen Xiaosheng et al. Changchun: Jilin Publishing Group Co., Ltd., 2013: 341.

[30] Coeckelbergh M. Moral appearances: Emotions, robots, and human morality[J]. Ethics and Information Technology, 2010, 3: 236.

[31] Dai Yibin. Research Approaches to Artificial Intelligence Ethics[J]. Social Sciences, 2023,(07):45-52.

[32] Zheng Huangjie. Trustworthy Artificial Intelligence: Governance Foundations for AIGC Amid Technological Ethical Risks [J/OL]. Science and Technology Progress and Countermeasures, 1-11 [2025-09-04]. https://link.cnki.net/urlid/42.1224.G3.20241101.1042.002.

[33] Dai Yibin. Research Approaches to Artificial Intelligence Ethics [J]. Social Sciences, 2023, (07): 45-52.

[34] Du Yanyong. Ethical Review of Artificial Intelligence: Current Status, Challenges and Solutions [J]. Journal of Donghua University (Social Sciences Edition), 2024, 24(02): 32-39.

[35] China Academy of Information and Communications Technology. Research Report on Ethical Governance of Artificial Intelligence (2023) [EB/OL]. (2023-12-26) [2024-12-26]. https://mp.weixin.qq.com/s/EUri4EgJGR9iSTuxMsRmzA.

[36] In June 2021, the World Health Organization (WHO) released the guidelines Ethics and Governance of Artificial Intelligence for Health.On November 24, 2021, the United Nations Educational, Scientific and Cultural Organization (UNESCO) adopted the Recommendation on the Ethics of Artificial Intelligence.On May 25, 2023, the United Nations (UN) issued Our Common Agenda.On October 26, 2023, the United Nations High-Level Advisory Body on Artificial Intelligence was established and released the interim report Human-Centered Artificial Intelligence Governance.

[37] In June 2018, the G7 Summit adopted the Shared Vision for the Future of Artificial Intelligence.In June 2023, the World Internet Conference held discussions under the theme "The AI Era: Building a Digital World of Exchange, Mutual Learning and Inclusiveness".In August 2023, leaders at the 15th BRICS Summit agreed to launch the work of the Artificial Intelligence Research Group as soon as possible, promote governance frameworks and standards with broad consensus, and continuously

ISSN: 2710-0170

enhance the safety, reliability, controllability and fairness of artificial intelligence technologies.In September 2023, the Group of Twenty (G20) released the G20 New Delhi Leaders' Declaration.

[38] In October 2016, the White House of the United States released the National Artificial Intelligence Research and Development Strategic Plan.In January 2020, the White House Office of Science and Technology Policy (OSTP) issued the Guidelines for Regulation of Artificial Intelligence Applications.In May 2022, the Biden administration established the National Artificial Intelligence Advisory Committee, which works on issues such as AI safety and prevention and mitigation of bias, to promote the development of responsible and inclusive artificial intelligence technologies.In October 2022, the White House of the United States released the Blueprint for an AI Bill of Rights.On October 30, 2023, U.S. President Joe Biden signed and issued the executive order Safe, Secure, and Trustworthy Artificial Intelligence.In February 2020, the U.S. Department of Defense promulgated the Ethical Principles for Artificial Intelligence.In May 2020, the U.S. Intelligence Community released the Artificial Intelligence Ethical Principles for the U.S. Intelligence Community.In July 2020, the Office of the Director of National Intelligence (ODNI) of the United States issued the Ethical Guidelines for Artificial Intelligence in the Intelligence Community.In January 2023, the National Institute of Standards and Technology (NIST) under the U.S. Department of Commerce released the Artificial Intelligence Risk Management Framework.In December 2023, the U.S. Government Accountability Office (GAO) released a report, presenting an assessment of the responsible use of artificial intelligence by 23 federal agencies.

[39] In April 2019, the High-Level Expert Group on Artificial Intelligence of the European Union (EU) released the Ethics Guidelines for Trustworthy Artificial Intelligence.On August 1, 2024, the EU Artificial Intelligence Act officially entered into force.

[40] In July 2017, the State Council of the People's Republic of China issued the Development Plan for the Next Generation of Artificial Intelligence.In 2018, the White Paper on Artificial Intelligence Standardization was released at the inaugural meeting of the National Overall Group for Artificial Intelligence Standardization and the Expert Advisory Group.In February 2019, the National Professional Committee for Next-Generation Artificial Intelligence Governance was established.In June 2019, the National Professional Committee for Next-Generation Artificial Intelligence Governance released the Governance Principles for the Next Generation of Artificial Intelligence – Developing Responsible Artificial Intelligence.In 2020, five departments including the Standardization Administration of China (SAC) promulgated the Guidelines for the Construction of the National Standard System for the Next Generation of Artificial Intelligence.In September 2021, the National Professional Committee for Next-Generation Artificial Intelligence Governance issued the Ethical Norms for the Next Generation of Artificial Intelligence.In March 2022, the General Office of the Communist Party of China (CPC) Central Committee and the General Office of the State Council jointly issued the Opinions on Strengthening the Governance of Science and Technology Ethics.In July 2023, seven departments including the Cyberspace Administration of China (CAC) released the Interim Measures for the Management of Generative Artificial Intelligence Services.

[41] Jia Kai, Xue Lan. Global Comparisons and Chinese Practices in Addressing Ethical Issues and Safety Risks in Artificial Intelligence Governance [J]. Public Administration Review, 2021, 3(01): 122-134.

[42] Zhao Jingwu. Pathways for Embedding Technological Ethics into Artificial Intelligence Governance Systems: The Case of Autonomous Driving Applications [J]. Rule of Law Society, 2024, (05): 16-28.

[43] JonathanDancy,EthicswithoutPrinciples,Oxford：ClarendonPress,2004,p.73

[44] Lei Haoran. On Agile Governance of Generative Artificial Intelligence Technology Ethics [J]. Journal of North China University of Water Resources and Electric Power (Social Sciences Edition), 2025, 41(01):50-58.

[45] Zhao Jingwu. Pathways for Embedding Technological Ethics into AI Governance Systems: An Examination of Autonomous Driving Application Scenarios [J]. Rule of Law Society, 2024, (05): 16-28.

[46] Ministry of Science and Technology of the People's Republic of China. Ethical Guidelines for New Generation Artificial Intelligence [EB/OL]. (2021-9-26) [2025-8-26]. https:// www. most . gov.cn/kjbgz/202109/t20210926_177063.html.

ISSN: 2710-0170

[47] He Miao, Chu Xingbo. Ethical Risks and Regulatory Choices in Generative Artificial Intelligence [J]. Studies in Dialectics of Nature, 2024, 40(11): 94-101.

[48] Xiao Hongjun, Zhang Lili. Theoretical Deconstruction of Ethical Deviance in Large Models and Governance Innovation [J]. Research on Financial and Economic Issues, 2024, (05): 15-32.

[49] He Miao, Chu Xingbo. Ethical Risks and Normative Choices in GenerativeArtificial Intelligence [J]. Studies in Dialectics of Nature, 2024, 40(11): 94-101.

[50] Zheng Huangjie. Trustworthy Artificial Intelligence: Governance Foundations for AIGC Amid Technological Ethical Risks [J/OL]. Science and Technology Progress and Policy, 1-11 [2025-09-04]. https://link.cnki.net/urlid/42.1224.G3.20241101.1042.002.

[51] Xiao Hongjun, Zhang Lili. Theoretical Deconstruction of Ethical Deviance in Large Models and Governance Innovation [J]. Research on Financial and Economic Issues, 2024, (05): 15-32.

[52] Zheng Huangjie. Trustworthy Artificial Intelligence: Governance Foundations for AIGC Amid Technological Ethical Risks [J/OL]. Science and Technology Progress and Policy, 1-11 [2025-09-04]. https://link.cnki.net/urlid/42.1224.G3.20241101.1042.002.

[53] Zhao Jingwu. Pathways for Embedding Technological Ethics into AI Governance Systems: The Case of Autonomous Driving Applications [J]. Rule of Law Society, 2024, (05): 16-28.

[54] Du, Yanyong. Ethical Review of Artificial Intelligence: Current Status, Challenges and Solutions [J]. Journal of Donghua University (Social Sciences Edition), 2024, 24(02): 32-39.

[55] Zheng, Huangjie. Trustworthy Artificial Intelligence: Governance Foundations for AIGC Amid Technological Ethical Risks [J/OL]. Science and Technology Progress and Countermeasures, 1-11 [2025-09-04]. https://link.cnki.net/urlid/42.1224.G3.20241101.1042.002.

[56] Central People's Government of the People's Republic of China. Opinions of the General Office of the CPC Central Committee and the General Office of the State Council on Strengthening the Governance of Science and Technology Ethics [EB/OL]. (2022-03-20) [2025-03-11]. https://www.gov.cn/gongbao/content/2022/content_5683838.htm.

[57] Du, Yanyong. Ethical Review of Artificial Intelligence: Current Status, Challenges and Solutions [J]. Journal of Donghua University (Social Sciences Edition), 2024, 24(02): 32-39.

[58] Xuefang Xie, Chen Qu. Value Alignment: Research on AI Cultural and Technological Ethical Risks and Pathways for Precise Co-governance in the AIGC Era [J]. Journal of Lanzhou University (Social Sciences Edition), 2024, 52(03): 147-156.

[59] Central People's Government of the People's Republic of China. Opinions of the General Office of the CPC Central Committee and the General Office of the State Council on Strengthening the Governance of Science and Technology Ethics [EB/OL]. (2022-03-20) [2025-03-11]. https:// www.gov.cn/gongbao/content/2022/content_5683838.htm.

[60] Meng Tianguang. Introduction to Intelligent Governance: AI-Driven Modernisation of Governance [M]. Beijing: Tsinghua University Press, 2023: 11.

[61] Xiao Hongjun, Zhang Lili. Theoretical Deconstruction of Ethical Deviance in Large Models and Governance Innovation [J]. Research on Financial and Economic Issues, 2024, (05): 15-32.

[62] Song Hualin. Constructing Ethical Norms for Artificial Intelligence from a Rule of Law Perspective [J]. Digital Rule of Law, 2023, (06): 1-9.

[63] Liu Zhiyi, Liang Zheng, Zheng Yejie. Black Mirror and Order: Artificial Intelligence Ethics and Governance in a Digital-Intelligent Risk Society [M]. Beijing: Tsinghua University Press, 2022.

[64] Wang Tianmin, Zheng Lili. Risks and Countermeasures of ChatGPT Empowering Ideological and Moral Education [J]. Studies in Dialectics of Nature, 2023, 39(08): 131-137.

[65] He Miao, Chu Xingbo. Ethical Risks and Normative Choices of Generative Artificial Intelligence [J]. Studies in Dialectics of Nature, 2024, 40(11): 94-101.

[66] Zhi, Zhenfeng; Liu, Jiakun. Ethics First: Governance Strategies for Generative Artificial Intelligence [J]. Yunnan Social Sciences, 2024, (04): 60-71.

[67] Meng, Tianguang. Introduction to Intelligent Governance: AI-Driven Modernisation of Governance [M]. Beijing: Tsinghua University Press, 2023: 11.

ISSN: 2710-0170

[68] China Academy of Information and Communications Technology. Research Report on Ethical Governance of Artificial Intelligence (2023) [EB/OL]. (2023-12-26) [2025-06-21]. https://mp.weixin.qq.com/s/EUri4EgJGR9iSTuxMsRmzA.

[69] Xuefang Xie, Chen Qu. Value Alignment: Research on AI Cultural-Technological Ethical Risks and Pathways for Precise Co-governance in the AIGC Era [J]. Journal of Lanzhou University (Social Sciences Edition), 2024, 52(03): 147-156.

[70] Hualin Song. Constructing Ethical Norms for Artificial Intelligence from a Rule of Law Perspective [J]. Digital Rule of Law, 2023, (06): 1-9.

[71] Central People's Government of the People's Republic of China. Opinions of the General Office of the CPC Central Committee and the General Office of the State Council on Strengthening the Governance of Science and Technology Ethics [EB/OL]. (2022-03-20) [2025-03-11]. https://www.gov.cn/gongbao/content/2022/content_5683838.htm.

[72] Zhi, Zhenfeng; Liu, Jiakun. Ethics First: Governance Strategies for Generative Artificial Intelligence [J]. Yunnan Social Sciences, 2024, (04): 60-71.

[73] Research Institute of Party History and Literature of the CPC Central Committee. Selected Statements by Xi Jinping on Preventing Risks and Challenges and Responding to Emergencies [M]. Beijing: Central Literature Publishing House, 2020: 79.

First Author: Zhao Yujian (born 1991), female, native of Beijing, currently pursuing doctoral studies in fundamental theoretical jurisprudence. E-mail: 1670859162@qq.com.