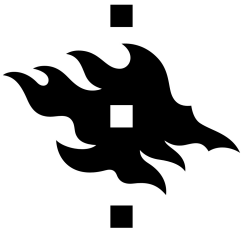


HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Comparative Analysis in AI Ethics: A Computational Study of Academic Discourses in China and the EU

Master's Programme in Linguistic Diversity and Digital Humanities
Digital Humanities
Master's thesis

Author:
Nan Jiang

Supervisor(s):
Professor in Digital Humanities (Human Sciences –Computing Interaction) Eetu Makela
University lecturer Anna-Mari Wallenberg
Professor of Digital Humanities Mikko Tolonen

2025-04-23
Helsinki

Faculty: Faculty of Arts

Degree programme: Linguistic Diversity and Digital Humanities

Study track: Digital Humanities

Author: Nan Jiang

Title:
Comparative Analysis in AI Ethics: A Computational Study of Academic Discourses in China and the EU

Level: Master's degree

Month and year: April 2025

Number of pages: 46

Keywords:
Artificial intelligence, AI ethics, China, European Union, Academic discourse, Large language models, Zero-shot classification, Regional comparison, Subject comparison, Digital humanities, Interdisciplinarity

Supervisor or supervisors:
Professor in Digital Humanities (Human Sciences –Computing Interaction) Eetu Makela
University lecturer Anna-Mari Wallenberg
Professor of Digital Humanities Mikko Tolonen

Where deposited: University of Helsinki digital repository (Helda)

Abstract: This research aims to study and compare the academic discourses of AI ethics in China and the EU regarding issues, mitigation proposals, and practitioners. It also seeks to explain the potential regional differences, temporal changes, and varying perspectives between social science and science subjects within the conventions of politics, culture, and stakeholders in AI. A total of 1442 journal articles on AI ethics, published between 2015 and 2024, were retrieved and collected, with 844 authored by researchers affiliated with Chinese institutions from the China National Knowledge Infrastructure (CNKI) and 598 authored by researchers affiliated with EU institutions from the Web of Science Core Collection. An advanced computational approach is employed to process and analyse the large-scale corpus, utilising the GPT-4o-mini model with a zero-shot classification method. This research provides a systematic overview and comparison of the academic discussion surrounding AI ethics in China and the EU. Furthermore, it offers insights for developing a more ethical AI world based on the patterns and differences identified through the comparative analysis between the two regions.

Contents

Abbreviations

AI	Artificial Intelligence
AGI	Artificial General Intelligence
CNKI	China National Knowledge Infrastructure
CoP	Community of Practice
CSCD	Chinese Science Citation Database
CSSCI	Chinese Social Sciences Citation Index
EU	European Union
GDPR	General Data Protection Regulation
LLMs	Large Language Models
SCI-Expanded	Science Citation Index Expanded
SSCI	Social Sciences Citation Index

1 Introduction

Artificial Intelligence (AI) has developed rapidly, and AI-related products have become embedded in everyday life due to their widespread functionality. However, it also raises many social issues(Stanford Institute for Human-Centered Artificial Intelligence (HAI), 2025). AI ethics is the subject that addresses these risks and the mitigation strategies to create a more ethical AI era(Weidinger et al., 2021).

Academia in different regions engages in varied discussions about AI ethics and shifts its focus across different phases of AI development. However, little research provides a systematic overview of how scholars from different regions differ in their discussions of AI ethics and how these discussions evolve over time (Bakiner, 2023; Zhu, 2024).Besides, China, as one of the major players in AI, has a significant impact on the global landscape of AI ethics. However, the manner in which Chinese research addresses AI ethics is not clearly understood in the West, particularly since some academic papers are published internally in Chinese. Meanwhile, the EU, as a leader in establishing supervision and regulation on technological issues, acts quickly and effectively concerning AI ethics(Floridi et al., 2018). While some studies have analysed the characteristics of AI ethics in China and the EU, a systematic comparison of academic discourses has yet to be conducted. Furthermore, research indicates that China and the EU have disparities in policies, culture, and practitioners when confronting topics related to AI(Cai & Yin, 2025; Fung & Etienne, 2023; Qiao-Franco & Zhu, 2022; Radanliev, 2025; Roberts et al., 2023; Ryan et al., 2021), but these factors have not been mapped to the complex academic discussions about AI ethics in these two regions. This research will explore, compare, and explain how AI ethics is discussed in China and the EU academia.

Another noticeable feature of AI is that, as a field highly dependent on technological development and intricacies, the attitudes and opinions of science scholars are relatively more crucial than in other areas of ethics, where social science scholars are the primary participants in debate(Cath et al., 2018). Hence, it is important to compare the opinions from different subjects to develop a more comprehensive understanding of AI ethics.

This study conducts a comparative analysis of academic discussions on AI ethics, examining issues, mitigation proposals, and practitioners in China and the EU across different periods and subjects to address the research gaps identified earlier. It analyses a substantial number of research papers from the two regions (China and the EU), two subjects (social science and science), and two time periods (2015-2029 and 2020-2024). Furthermore, it analyses how regional conventions in politics, culture, and practitioner involvement influence the observed differences and reveal intriguing patterns. This approach aims to enhance the understanding of global AI ethics debates to foster more diverse and inclusive collaborations and tailor measures to local conditions, which are essential for improving the global state of AI ethics.

Last but not least, the traditional methods previously used often involve manual annotation for all papers,

which limits the dataset size and hinders efficiency (Holzinger et al., 2019). Hence, another significant breakthrough of this research is using computational methods to effectively analyse large-scale, long-text data, achieving a more comprehensive outcome efficiently. It is also possible to create a dynamic supervision system for the academic discourse on AI ethics. This research examined 844 Chinese academic articles from the China National Knowledge Infrastructure(CNKI) and 598 EU academic articles in English from the Web of Science Core Collection[1]. The GPT-4o-mini, a Large Language Model (LLM), is utilised for zero-shot classification with prompts to determine whether a list of given topics on AI ethics is discussed in each paper. The results are then calibrated against results from the manual annotation of 78 sampled papers, which are used for the final comparison of regions, subjects, and periods. In summary, this research builds a new corpus consisting of 1442 academic papers on AI ethics from China and the EU between 2015 and 2024. It analyses the topic density using the latest LLM within a nuanced methodological framework that includes computational methods, manual annotation, and calibrations.

In conclusion, this research utilises advanced computational methods to analyse a large academic dataset on AI ethics from China and the EU, comparing them across regions, periods, and subjects. Additionally, it elucidates the differences by employing conventions from politics, culture, and practitioners to provide insights into understanding AI ethics academic discourses in various contexts and how to foster a more ethical AI world by integrating best practices from different regions.

2 Related Work in AI Ethics

2.1 Topics of AI Ethics in China and the EU

Previous research in China(吴汉东, 2017; 李修全, 2017; 段伟文, 2017; 赵志耘 et al., 2021; 马长山, 2018) and the EU (Fjeld et al., 2020; Giarmoleo et al., 2024; Hagendorff, 2020; Weidinger et al., 2021, 2022) shows that AI ethics is mainly discussed from the perspectives of issues, mitigation proposals, and practitioners. In these three aspects, there are similarities and differences in the specific topics academia discusses between the two regions.

China and the EU face shared issues in AI ethics, such as protecting human rights, ensuring fairness, promoting accessibility, safeguarding privacy, and enhancing transparency. They assess whether AI supports fundamental human rights like freedom, privacy, equality, and freedom of speech. Furthermore, both regions confront risks associated with technology, such as security, ensuring AI systems protect individuals and infrastructure from attacks and data breaches, and autonomy, emphasising respect for individual decision-making and defining the roles of AI agents. Additionally, they consider societal impacts, including potential job losses due to automation, and misinformation from AI-generated content that threatens public trust. Environmental concerns also emerge in these two regions, as AI development

[1]In this thesis, references to China and the EU as regions for collected journal articles denote the geographic affiliation of authors' institutions, as determined by the Web of Science "Countries/Regions" filter for EU papers and CNKI for Chinese papers. This definition applies throughout the document.

consumes significant energy and affects ecosystems. These issues shape the ethical landscape of AI risks in both regions.

China and the EU also share common proposals for mitigating AI challenges. Both regions implement legislation with frameworks for the ethical use of AI, providing enforceable guidelines, and they establish ethical and moral standards to guide AI ethics practices simultaneously without mandates. Industries are required to seek technical solutions to optimise algorithms and models, and corporations are asked to take responsibility for self-regulation in development, promotion, and application. Both regions advocate for reliable and responsible AI use to ensure accountability. They also believe that public understanding is important and should be enhanced through education about AI algorithms to reduce misconceptions. Finally, international cooperation is emphasized to foster collaboration on AI ethics, addressing global challenges through communication and mutual support. These proposals collectively advance ethical AI practices.

The governance of AI ethics involves various participants, each with unique roles. Both China and the EU include five stakeholders: Government institutions, such as legislative bodies, formulate AI ethics policies for regulatory oversight. Scholars and experts, like ethicists, offer theoretical support and expert advice to ground ethical frameworks in research. Corporations and enterprises engaged in AI technology development play a crucial role in applying ethical practices. The general public, including users and affected individuals, voices concerns and advocates for the fair and ethical use of AI. Lastly, international organisations, like the United Nations, promote multilateral cooperation on AI ethics and global standardization. Together, these participants create a collaborative ecosystem to tackle AI ethics challenges.

Aside from the common topics introduced above, China and the EU have their own concerns as well. When examining the risks of AI discussed by 吴汉东 (2017), 李修全 (2017), 段伟文 (2017), 赵志耘 et al. (2021), and 马长山 (2018), Chinese academia addresses numerous social risks, such as algorithm misuse, which refers to the inappropriate, unethical, or harmful application of AI algorithms, whether intentional or unintentional, leading to negative consequences for individuals, groups, or society, such as surveillance. Uncontrollability questions whether AI systems exhibit unpredictable or uncontrollable behaviour that transcends human understanding or control. To address these concerns, Chinese scholars propose several mitigation strategies in addition to the shared solutions: in research development, by focusing on human-centred values and promoting openness and sharing; in governance, by fostering harmony and cooperation and implementing long-term planning; and in usage and technological advancements, by advocating for consumers' careful use.

There are also some risks related to AI explicitly mentioned in the EU research(Fjeld et al., 2020; Giarmoleo et al., 2024; Hagendorff, 2020; Weidinger et al., 2021, 2022). First is sustainability, which refers to the responsible development and use of AI technologies to minimise negative environmental, social, and economic impacts. Second, malicious uses highlight the significant adverse effects that AI technologies could have on humans or society, including physical and psychological harm, as well as safety risks such as lethal autonomous weapons. Furthermore, diversity and inclusion emphasise ensuring that diversity is considered in the development, design, and application of AI technologies so that the needs and rights of diverse populations are respected and addressed. The approaches to addressing these issues from these studies also provide some solutions in addition to the common ones. The EU primarily

focuses on human oversight, intensified research funding, and ethical product design. Human oversight ensures that humans supervise and engage in the design, development, and application of AI systems, maintaining accountability and ethical alignment. Ethical product design emphasises incorporating ethical principles throughout the entire lifecycle of AI application development to ensure their use aligns with moral standards. Academic research, through theoretical and empirical studies in various AI-related fields, provides scientific evidence for policymaking and technology practices, grounding solutions in robust data.

Therefore, there are many commonalities and some potential distinctions in academic discourses on AI ethics between these two entities. However, these findings are primarily based on limited close reading, and the prioritisation of topics can not be fully justified in this manner. Hence, it is necessary to conduct a systematic analysis of a much larger dataset to validate the differences in the content and priority of AI ethics topics in China and EU academia. Additionally, regional conventions can help to understand and explain the regional divergences in AI ethics discussions between China and the EU, which will be studied in the next section.

2.2 Regional Conventions Impacting Academic Discussion of AI Ethics in China and the EU

To explain the potential divergences in academic discussions of AI ethics across regions, it is necessary to investigate the different strategies and traditions of China and the EU in politics, culture, and practitioners regarding AI. These conventions provide opportunities to explore the underlying causes of regional differences.

2.2.1 AI Ethics Policies in China and the EU

The first part of the regional conventions involves the policies in the AI field proposed by government institutions. This section will introduce three components of these AI policies: aims, approaches, and beneficiaries, which enhance our understanding of the underlying political frameworks that academic discussions on AI ethics may navigate in specific ways across various regional contexts.

2.2.1.1 Aims in AI Governance

The various starting points at the national level will shape practices regarding AI ethics across different sectors, including academia, as they indicate an overall direction and establish bottom lines for most units in society. According to research from Roberts et al. (2023), the AI policies in China and the EU are both dedicated to promoting global AI leadership, improving social welfare, and enhancing economic performance while striving to mitigate potential risks as they formulate AI regulations.

Nonetheless, Roberts et al. (2023) also points out that some aspects of their objectives differ, as summarized in the table 2.1. The ignition of efforts in China aims to secure competitive advantages in technology, economics, and national security in response to the emergence of AlphaGo. In contrast, the EU primarily focuses on the negative social consequences of the widespread implementation of AI.

Furthermore, since the EU is not a sovereign country, military concerns can not be a shared aim among its member nations; however, China prioritizes using AI at the national level defence(Roberts et al., 2021). These differences imply that China prioritises an AI ethics strategy to enhance economic and military competitiveness, while the EU tends to view AI ethics more through the lens of human rights, with little emphasis on military considerations and less focus on the economy.

Table 2.1: Comparison of AI Governance Aims Between China and the EU

Aims	China	the EU
Origins of AI ethics	"Sputnik Moment" of AlphaGo's 2016 and 2017 victories over champion Go players ignite the effort in AI ethics	Widespread adoption of AI and a desire to control the potentially harmful impacts of use
Military considerations	To win new competitive advantages in national defence	All military-related decisions belong to each independent country

2.2.1.2 Approaches in AI Governance

As for measures to solve AI ethics issues compared in the table2.2, China prefers to tackle problems as they arise, while the EU tends to steer clear of risks beforehand(Roberts et al., 2023). By employing two distinct strategies, their incentive approaches target different parties. China has implemented a system to promote market innovation, while the EU focuses on safeguarding fundamental human rights(Sajduk & Dziwisz, 2024). For instance, the EU AI Act suggests establishing a risk-based classification system for AI applications(Runcan et al., 2025). This type of pre-intervention is uncommon in China. However, both lack some degree of public involvement, particularly in the oversight process(Roberts et al., 2023)

Table 2.2: Comparison of AI Governance Approaches between China and the EU

Approaches	China	the EU
Strategy	Ad hoc measures to curtail harms	Set ethical parameters and provide upfront support
Incentives	A system of innovation incentives for public and private actors	Focus on protecting fundamental rights
Oversight	Lack of supervision from general public	

The potential impact of approaches on academia is that Chinese academic discourse may adopt more incentive strategies to address AI issues while relatively sidelining potential risks. Meanwhile, the EU is likely to deprioritise the positive effects of the AI market, opting instead to focus on implementing preemptive measures aimed at preventing AI risks and protecting human rights. However, both parties overlook the voice of the public.

2.2.1.3 Beneficiary in AI Governance

Echoing the aims and approaches of each region, the focus of AI ethics beneficiaries in China primarily revolves around international competitiveness, particularly economic strength. Meanwhile, the EU prioritises protecting individuals' rights(Roberts et al., 2023). The leading beneficiary also contributes to shaping the academic discourse on AI ethics.

Taking fairness as an example from Roberts et al. (2023), aside from addressing political concerns regarding social stability, China emphasizes “consumers .” This suggests that China tends to place individuals in the market context rather than in civic life, which could stem from a priority on an economy-centred approach. Therefore, economic interests can be central to Chinese academia in discussions of AI ethics.

One crucial factor that leads the EU to prioritise human rights is its legal commitment to the Charter of Fundamental Rights of the EU(European Union, 2012), which serves as a foundational framework for human rights within the EU’s legislative system. This commitment significantly influences AI ethics topics, as the Charter mandates that all EU policies, including those governing AI, uphold fundamental rights such as non-discrimination, privacy, equality, and freedom of expression. In the context of AI ethics, this legal obligation shapes regulations like the EU AI Act(European Parliament and Council of the European Union, 2024) and the General Data Protection Regulation (GDPR)(European Parliament and Council of the European Union, 2016), ensuring that principles such as transparency, fairness, and inclusivity are embedded in AI development and deployment.

In China, there are no formal laws regarding AI ethics. According to Dong and Chen (2024) and Sheehan (2023), the most similar laws are related to data and privacy issues in the digital realm, such as the Cybersecurity Law(State Council of China, 2016), Data Security Law(National People’s Congress of China, 2021), and Personal Information Protection Law(Standing Committee of the National People’s Congress, 2021). However, these laws do not explicitly address AI issues and primarily prioritise state security and social stability in the context of the big data era. Additionally, the Civil Code(National People’s Congress, 2020) contains Tort Liability Rules that can hold AI developers accountable for harm, but it does not cover all cases in the AI sector concerning fundamental human rights protections. Regarding AI-specific regulations, there are non-mandated frameworks such as the Governance Principles for New Generation AI (Ministry of Science and Technology of China, 2019), Code of Ethics for the New Generation AI (Ministry of Science and Technology of China, 2021), and Interim Measures for the Management of Generative Artificial Intelligence Services (Cyberspace Administration of China, 2023). They emphasise constraining generative AI service providers regarding data control and information security, rather than protecting users’ fundamental rights from the outset. Again, these regulations are not enforced in cases related to AI ethics.

However, regarding beneficiaries in AI ethics, there is a social and state value in China that reflects the importance of human well-being, “harmony” and “harmonious society”. According to Cai and Yin (2025) and Han (2008), this concept emphasises the realisation of individual happiness and shared prosperity through social stability. It represents a dynamic balance and order among all stakeholders to achieve maximum collective interests with minimal conflicts. From the perspective of individuals, it does not mean everyone achieves the same level of interest, but rather that they feel their values and desires are attainable. Additionally, this perception prioritises social cohesion, believing that only under these conditions can the prosperity of society be realised. In the context of AI, government, technology companies, research institutions, and the public should work together to advance AI development. During this process, individual interests should be respected and protected to help build a sense of harmony among people, but the ultimate goal is to achieve the greatest collective development; thus, individual interests can yield to others when necessary. However, severe violations of public rights are not allowed if they influence the stability of the entire social system.

In conclusion, the hierarchy of beneficiaries, along with the legislation or social values that promote the practice, differs between China and the EU. This distinction will also influence how academia discusses AI ethics.

2.2.1.4 Fusion of AI Policies in China and the EU

Some cases indicate that the boundary of AI governance between the two entities has begun to blur, complicating academic debates on AI ethics in both regions.

First, it is widely accepted that the state's interests always take precedence over any other principles in China. Similarly, Roberts et al. (2023)mentioned that the EU's GDPR permits exceptions for consent and notification requirements in situations involving state obligations. Second, according to Cai and Yin (2025), Chinese AI ethics principles emphasise the importance of "Harmony, " a concept that asserts AI development should respect basic human rights and create a harmonious coexisting environment for humans and AI. Otherwise, the state of discord will not only hurt humans but also negatively impact the long-term development of AI. Therefore, the representation of the "harmonious society" goal of China in AI aligns with the emphasis on human rights in EU policies. Third, 贾婷 and 沈天添 (2023) highlighted the shift in global trends in AI ethics from moral considerations to practical implementations, moving from individual actions to collaborations among various stakeholders. Moreover, the global context, especially the shift towards deglobalization and possible restrictions on technology transfer, presents challenges for both regions(Sajduk & Dziwisz, 2024).

Therefore, when discussing AI ethics in academic discourse in China and the EU, the topics of national interests, the coexistence of AI and humans, specific strategies to address AI ethical issues, and reactions to the international context become complex and require further study.

In summary, the political considerations in the two regions exhibit significant differences in the aims, approaches, and beneficiaries of AI ethics policies, with some cases indicating a trend of fusion. These political factors will influence academic discussions about AI ethics between China and the EU in specific ways, which will be explored in this thesis.

2.2.2 Cultural Backgrounds in China and the EU

In addition to political factors, the region's historically rooted culture significantly influences academic discourse(Cath et al., 2018). Researchers are more likely to incorporate cultural considerations when illustrating contemporary topics(Triandis, 1993). Therefore, it is vital to study the cultural differences that may influence how academia develops its understanding of AI ethics in China and the EU.

2.2.2.1 Confucianism and Liberalism

The primary cultures influencing academic discussions on AI ethics in China and the EU are grounded in Confucianism and liberalism, respectively (Cai & Yin, 2025).

In China, Confucianism emphasizes the strong bond between individuals and society, highlighting their reciprocal impact. It also indicates that collective values may, when necessary, take precedence over

individual interests to achieve social harmony(Cai & Yin, 2025). This cultural perspective suggests that national interests are more important than civil rights, indicating that China will likely prioritise benefits for political entities over individual rights when addressing AI issues. Meanwhile, maximising societal benefits is at the core of Confucianism, so technology can be utilized relatively freely as long as it serves the interests of the nation as a whole, which also practically empowers tech companies in the AI sector. In contrast, liberalism centres on individual rights and freedom, viewing the individual as the fundamental unit of society(Cai & Yin, 2025). Therefore, in AI ethics, EU academia is likely to emphasise individual rights such as privacy, transparency, and contractual freedom, including the accountability of commercial companies.

Additionally, regarding individuals, Cai and Yin (2025) stated that Chinese culture prioritizes responsibility over freedom and obligation over rights. Therefore, China is likely to place more demands on the public rather than empower them to take charge, as seen in the EU. Furthermore, scholars bear a significant responsibility to serve society in the traditions of Confucianism. In this way, academia is expected to be more involved in the discussion of AI ethics.

2.2.2.2 Utopian and Dystopian Visions

The level of trust in government and technology will also influence how academia in China and the EU approaches AI ethics issues.

As Fung and Etienne (2023) mentioned, Chinese guidelines based on Confucian principles emphasize virtuous governance and social harmony. Consequently, Chinese approaches are primarily promotional and avoid conflicts because they assume a shared trust within society, and the civic sector believes that the government will effectively guide the development of society. As a result of this vision, the government and technology have gained more power in China, but the public, including academia, may relinquish some personal rights to achieve the overall goal of society. For instance, the government employs facial recognition technology to enhance public security, facilitated by AI technologies that monitor public spaces at the expense of individual privacy to achieve social order security(Radanliev, 2025).

In contrast, the EU framework is founded on essential Enlightenment principles of individual rights and safeguards against state abuses (Radanliev, 2025). In this context, the EU primarily employs prohibitive methods, as no one can be trusted without sufficient oversight. To foster users' trust, its ethical framework operates as a dialectical system that balances the needs of users with those of AI developers and service providers(Fung & Etienne, 2023). Radanliev (2025) provides an example in healthcare: the EU's AI Act mandates that AI models ensure transparency in their decision-making processes, enabling audits by healthcare professionals and regulators. Therefore, transparency and human oversight are crucial to discussions of AI ethics within the EU academia.

Regarding obligations, Fung and Etienne (2023) stated that when conducting regulations, China's utopian vision tends to combine strict deontological norms with more flexible guidelines that encourage individuals to take responsibility for fostering social harmony, such as virtue ethics and "self-discipline." In contrast, the EU is influenced by cyberpunk culture, adopting normative rules of deontology for tech companies instead of allowing the public to take on the responsibility, as is the case in China(Fung & Etienne, 2023). Therefore, EU academia tends to recognise users' scepticism about technology and private

companies, aiming to mitigate this negative perception through protective measures and regulations for large corporations.

In summary, Confucianism and China's utopian vision foster social governance practices that demonstrate greater trust in the government while increasing demands on individuals and scholarly groups. In contrast, liberalism and dystopian traditions in the West lead the EU to question and oversee the power of government and business sectors, thereby empowering the public. These cultural traditions are expected to illuminate some divergences in the academic discourse on AI ethics between China and the EU as well.

2.2.3 Community of Practice(CoP) of AI Ethics in China and the EU

Beyond the influence of politics and culture, practitioners in the field of AI play a crucial role, especially when academia discusses accountability surrounding AI ethical issues (Morley et al., 2020). Governments, private companies, the public, academia, and international organizations are the main actors in the realm of AI ethics. It is essential to explore their functions and relationships in the practices of China and the EU before analysing their roles in academic discourses.

2.2.3.1 Community of Practice(CoP) of AI Ethics in China

Within the realm of AI ethics, China has established a distinctive community of practice that includes three sectors: government, academia, and private companies (Qiao-Franco & Zhu, 2022).

First, the CoP in China is spearheaded by the government. The government can influence who is included or excluded, especially in large, nationally directed strategic technology initiatives programs(Qiao-Franco & Zhu, 2022). This quality of the community empowers the government at the highest level, which may lead to preferential treatment and fewer limitations regarding the AI employed by the government. For instance, Qiao-Franco and Zhu (2022) mentioned that government agencies might bypass data protection obligations when faced with situations concerning national security, such as the establishment of the Social Credit System.

However, the private sector is increasingly taking charge of establishing technical standards and norms for AI development and implementation in China, as it wields far greater influence in this era of AI (Qiao-Franco & Zhu, 2022). Therefore, the significance of private companies is growing in China, which may lead to increased discussions about AI companies in academia.

The roles of academia and the public are marginalised in the Chinese CoP. In academia, experts and researchers primarily collaborate with the government and businesses through informal communication methods and formal bureaucratic processes(Qiao-Franco & Zhu, 2022), to provide expertise for policymakers to develop and disseminate guidelines, but they do not have a say in the final decision-making. Moreover, the public's absence in this community highlights the overlooked position of ordinary people in AI regulations. This situation may lead to a decline in the discussion of topics related to these two parties in academic circles, including misinformation, the protection of human rights, long-term environmental harm, and funding for AI regulation studies. However, academia may wish to address this unfair situation by encouraging greater discourse in these areas.

Despite this imbalanced relationship, non-state actors' participation in China's CoP will continue. According to Qiao-Franco and Zhu (2022), there are two reasons. The first is that they can gain government resources and support through this cooperation. Second, when facing challenges from Western counterparts, Chinese companies collaborate with the government and researchers to address problems. At the same time, domestic ethical challenges compel these policymakers to develop and share guidelines collaboratively. Many institutions incorporate these three stakeholders to establish principles for AI ethics. Therefore, we will likely observe the increasing role of academia and private companies in Chinese academic discourses. In particular, academia may seek greater attention through internal discussions but will not completely disregard the viewpoints of other parties or compromise the comprehensiveness and representativeness of its research. Furthermore, it is unlikely for academia to sacrifice AI development to propose necessary precautions, considering the shared interests of CoP.

2.2.3.2 Leading Actors of AI Ethics in the EU

As stated by Ryan et al. (2021), the EU government leads in AI ethics practices; however, it also encounters risks of governmental exclusions, including deliberately obscuring the monitoring of political unrest and its possible repercussions. This scenario mirrors that of China. So the discussion of government will also be prioritized in EU academia.

Besides, European standards organizations are at the forefront of establishing standards(Roberts et al., 2023), so the EU risks undemocratic governance, where only elite voices are heard while numerous interest groups in society struggle to engage with these institutions. Moreover, users outside technology companies do not fully understand how AI utilizes their data and impacts them unconsciously (Ryan et al., 2021). This may give rise to a false sense of security for users of AI products, and the essential education of the public to improve AI literacy may be emphasised to address this problem.

In summary, the EU's leading actors in AI ethics face challenges such as excessive governmental power, an overrepresentation of elite discourse, and political disconnection among ordinary people due to a lack of AI literacy. Scholars may recognise that certain facts exist, so the discussion of government and organizations will not diminish, but mitigation proposals such as public education and human oversight may be discussed more frequently.

In conclusion, the main practitioners in AI ethics exhibit both differences and commonalities between China and the EU. Additionally, the relational dynamics and hierarchy differ across these two regions. This research will explore how academia is shaped by, or reacts to, these two communities of practice in AI ethics discussions.

2.3 Subject Features in AI Ethics Discourses

Aside from regional and temporal differences, the subject also influences the AI discussion, as the technological aspect is crucial in the field of AI. Generally, social science researchers aim to answer questions about how humans should behave. In contrast, scholars in science fields focus on understanding the capabilities of computer systems and technology(Stenseke, 2022).

As Cai and Yin (2025) said, the ethical landscape of AI in social sciences research is becoming increasingly politicised. Meanwhile, many studies show that AI researchers in the science field are well-equipped to identify new risks and develop technical solutions because of their proximity to the technology itself (Zhang et al., 2021; 李修全, 2017; 赵志耘 et al., 2021). Another reason for the increased discourse power of science researchers in AI is that AI-related researchers gain significant global attention. As Stanford Institute for Human-Centered Artificial Intelligence (HAI) (2025) noted, AI's growing importance is reflected in major scientific awards, including two Nobel Prizes and the Turing Award, which honour groundbreaking contributions in AI.

Moreover, there is a growing trend in merging social science with science research, particularly in interdisciplinary governance concerning applied AI ethics (Stanford Institute for Human-Centered Artificial Intelligence (HAI), 2025). Approaches such as embedded ethics, participatory AI design, and co-regulation with policymakers are increasingly popular.

However, statements from researchers in the field of science regarding AI ethics lack thorough investigation. Meanwhile, the opinions of social science scholars are also worth further study, as they have evolved in recent years alongside the rapid advancements in AI. This research will help identify the characteristics of each group about the discussion on AI ethics, particularly focusing on how science scholars understand and contribute to this area through their technical expertise. Additionally, it offers an opportunity for interdisciplinary and transdisciplinary dialogue, exploring how researchers in digital humanities can serve as a bridge between social science epistemologies and computer science technologies in AI ethics.

2.4 Previous Meta-Analysis about AI Ethics

Some systematic literature reviews offer a comprehensive assessment of the overall distribution and trends of AI ethics topics in academia(Cai & Yin, 2025; Hagendorff, 2020; Jobin et al., 2019; Tran et al., 2019). However, much of the data is outdated, creating a significant gap given the recent surge in AI applications. These reviews also do not cover dedicated comparisons between China and the EU regarding issues, mitigation proposals, and practitioners. Furthermore, Giarmoleo et al. (2024) shows that the conventions underlying these distinctions, such as politics and culture, are rarely explored. Nonetheless, the conventional factors influencing the focus of academic discussion on AI ethics are worth studying to gain a better understanding of why China and the EU discuss AI ethics in a specific manner and how to identify common points while acknowledging differences.

In one of the previous studies of Zhu (2024), a clear structure for systematically analyzing academic discourse in AI ethics was developed. This study analysed 328 publications from 2011 to 2020, sourced from CNKI. The search term "人工智能伦理 (AI ethics)" was employed. All articles were manually coded and labelled across six categories. The final findings explore the distribution of academic articles across various categories, including issues, proposals, guidelines, gender, and affiliations related to AI ethics research.

This research adheres to the structure of Zhu (2024) while enhancing it in five key areas to address gaps in previous studies. First, to improve the robustness and scope of the AI ethics research, the dataset will

be updated and expanded to cover the most recent decade, from 2015 to 2024, ensuring a comprehensive temporal analysis. Second, a dataset from the EU will be included to facilitate comparisons between China and the EU, enabling a deeper understanding of regional differences in AI ethics discourse. Third, scholars from science faculties, including computer scientists, engineers, mathematicians, and other science disciplines, will be incorporated to capture their perspectives, enriching the interdisciplinary dialogue. Fourth, an analysis of potential relations between academic discourse in AI ethics and regional conventions will be added to explore how different contexts shape AI ethical frameworks. More importantly, to support these objectives, an advanced computational method leveraging large language models (LLMs) will be utilised to analyse the expanded dataset, enhancing the coverage and depth of the findings.

3 Research Questions

Based on the literature review and the identified research gaps, the main research questions are: What are the focuses and priorities of topics in AI ethics in Chinese and EU academia from 2015 to 2024 regarding issues, mitigation proposals, and practitioners? What possible explanations can be drawn from regional conventions? These two questions address the differences in the academic discourse of AI ethics between China and the EU over the last decade and provide explanations from the perspective of politics, culture, and practitioners.

There are two additional sub-questions: 1. What are the temporary changes from 2015 to 2024? This question examines whether the differences in the academic discourses of AI ethics evolve over time due to varying stages of AI development. 2. What are the differences between social science and science scholars? This question facilitates a comparison of the perspectives of scholars from different subjects on AI ethics.

4 Methodology

This section covers the overall structure and specifics of employing computational methods. It begins with selecting topics related to issues, mitigation proposals, and participants in AI ethics from the literature review, which will be examined in the corpus. Next, raw data will be collected from two well-known archives of academic papers in China and the EU, following a defined retrieval method. Some papers will be sampled for manual annotation at the same time. Subsequently, one LLM model will be selected to analyze the corpus with a prompt asking whether the defined topics are discussed in each paper, and its results will be evaluated against the annotations. After adjusting the original modelling results, normalization is conducted to generate “Topic Density, ” which is the frequency of discussion of each topic per paper. The final analysis is conducted using this metric.

4.1 AI Ethics Topics to Examine

After reviewing previous AI ethics research by Cath et al. (2018), Floridi et al. (2018), Giarmoleo et al. (2024), Hagendorff (2020), Jobin et al. (2019), Mittelstadt (2019), and Ryan et al. (2021), a list of key topics related to AI ethics in academia is generated. This list includes shared and region-specific topics from China and the EU, classified into three categories: issues, mitigation proposals, and practitioners. The model will evaluate whether these topics are addressed in the collected academic papers.

Here are the topics in each category, along with their definitions.

4.1.1 Issues in AI Ethics

1. "Fairness": The degree to which AI technology and products treat all individuals equitably, avoiding discrimination or bias based on race, gender, creed, or social status class.
2. "Inclusivity": Ensuring that diversity is considered in the development, design, and application of AI technologies so that the needs and rights of diverse populations are respected and addressed.
3. "Accessibility": Whether AI technologies are widely accessible and affordable for underprivileged groups or regions.
4. "Transparency issues": The degree to which AI algorithms and decision-making processes are transparent enough to be understood and monitored by the public or regulatory bodies.
5. "Privacy": Ensuring the protection of personal data during the collection, storage, and use of AI applications.
6. "Human rights": Whether AI upholds fundamental human rights such as freedom, privacy, equality, and liberty of speech.
7. "Security": The reliability and resilience of AI systems in preventing malicious attacks, system failures, and data breaches, while keeping individuals, societies, and infrastructures safe from the risks posed by AI.
8. "Severe harms": The significant adverse impacts AI technologies could have on humans or society include physical and psychological harm, along with safety risks (e.g., lethal autonomous weapons).
9. "Autonomy and agency": Whether AI respects individual decision-making rights and the potential risks of autonomous and uncontrollable behaviour in AI systems, as well as how to address the roles and rights of AI agents.
10. "Uncontrollability": Whether AI systems demonstrate unpredictable or uncontrollable behaviour that transcends human understanding or control.
11. "Misinformation ": The risk posed by AI-generated or propagated false or misleading information to societal discourse and democratic systems.
12. "Unemployment": The potential negative impacts of AI on the job market include job losses resulting from automation and increased social inequality within the economic context.

13. “Environmental harms”: AI technologies’ negative effects on natural resources and ecosystems during their development and usage, as operating them requires significant energy resources.

4.1.2 Solution Proposals in AI Ethics

1. “Legislation”: Establishing legal and policy frameworks to regulate the ethical use of AI.
2. “Ethical and moral standards”: Establishing ethical guidelines and behavioural principles for AI technologies to effectively direct their development and application without mandates.
3. “Academic research”: Theoretical and empirical studies in various AI-related fields provide scientific evidence for policymaking and technology practices.
4. “Corporate responsibility”: AI companies adopting self-regulatory measures to address ethical issues in their development, promotion, and application processes of technology.
5. “Trustworthy and responsible use”: Conduct audits to ensure AI technologies are developed and applied reliably and trustworthily, with accountable entities overseeing the process.
6. “Human oversight”: Humans oversee and participate in the design, development, and application of AI systems.
7. “Technical solutions”: Utilising technical methods such as algorithm optimisation and model enhancements to address ethical challenges in AI technologies.
8. “Ethical product design”: Incorporating ethical principles throughout the entire lifecycle of AI application development ensures that their use aligns with moral standards.
9. “Algorithmic literacy”: Educating the public and stakeholders about how AI algorithms operate and affect their lives, thereby reducing opacity and misconceptions.
10. “Public education”: Raising public awareness of AI technologies through education and helping society recognise and advocate for solutions to ethical issues.
11. “International cooperation”: Collaborating among countries on AI ethics to tackle global challenges through communication and mutual support.

4.1.3 Participants in AI Ethics

1. “Government institutions”: State-level decision-making bodies, such as government departments or legislative institutions, are responsible for formulating policies related to AI ethics.
2. “Scholars and experts”: Academic professionals and technical experts, such as ethicists or those in related fields, provide theoretical support and expertise.
3. “Corporations and enterprises”: Private sector organisations involved in developing and commercialising AI technologies.

4. “General public”: The end users and individuals affected by AI technologies, including everyday citizens, community organisations, and consumers.
5. “International organisations”: Cross-border institutions, such as the United Nations, foster multilateral cooperation on AI ethics issues and advance global standardisation.

4.2 Raw Data Collection

4.2.1 Database

Chinese papers are collected in the China National Knowledge Infrastructure (CNKI) following the replicable research of Zhu (2024)[1]. CNKI, the largest global provider of Chinese academic information, offers a comprehensive range of digital resources, including academic journals, dissertations, e-books, newspapers, and conference papers, accessible to users in over 100 countries and regions beyond China, including more than 1,000 institutions(CNKI, n.d.).

The EU papers come from the Web of Science Core Collection, a leading global citation database[2]. This collection encompasses records of articles published in both open-access and traditional impactful journals, as well as conference proceedings and books. It includes ten indices that gather information from thousands of scholarly journals, books, book series, and conferences (Clarivate, n.d.-b).

4.2.2 Indices

To ensure the quality of the corpus, only articles referenced from reputable indices will be selected.

For Chinese papers, social science articles are sourced from journals listed in the Chinese Social Sciences Citation Index(CSSCI). This citation database specialises in Chinese social sciences journals, assessing their quality and impact in fields such as economics, law, education, and political science (University, n.d.). The journals in CSSCI are widely recognised as prestigious within the social sciences community. Science articles originate from the Chinese Science Citation Database(CSCD), a leading Chinese citation index for STEM disciplines. According to Chinese Academy of Sciences (2023), it is often viewed as the Chinese equivalent of the Science Citation Index (SCI).

For the EU, social science papers derive from the Social Sciences Citation Index(SSCI). It encompasses a broad range of social science disciplines and is an essential resource for tracking citations and measuring the impact of social science research (Clarivate, n.d.-b). Science papers are sourced from the Science Citation Index Expanded(SCI-Expanded), a premier global citation index for natural and applied sciences and an enhanced version of the original Science Citation Index (Clarivate, n.d.-a).

4.2.3 Search Terms

China: 人工智能 * (道德 + 伦理)[3]

[1] All the papers in CNKI are written in Chinese by scholars in Chinese institutions.

[2] All the collected EU papers are written in English.

[3] According to the syntax requirements of Boolean logic operations, spaces must be added on both sides of the logical operators to connect them with the search terms, and parentheses must be in half-width English format. Otherwise, the search's outcome will be different.

According to CNKI (n.d.), this search term returns papers that contain (人工智能 AND 道德) OR (人工智能 AND 伦理) in the search field. The meanings of "道德" and "伦理" in China are similar but emphasise different aspects. As Hansen (1992) said, "道德" often highlights personal conduct and internalised beliefs, while "伦理" is more focused on collective or institutional standards of behaviour. To achieve comprehensive coverage of "AI ethics" in Chinese, I added "道德" to the original search term "人工智能 伦理" used in the replicable research of Zhu (2024).

The EU: artificial intelligence ethics

According to Clarivate (2025a), the results of this research term include all documents that contain "artificial" AND "intelligence" AND "ethics" in a Topic search[4].

4.2.4 Retrieval Methods

4.2.4.1 China

The "Subject" search item is used to find Chinese Papers in CNKI. This search method employs subject fields recommended by CNKI, which encompass all relevant topics in an article. It ensures that search results are comprehensive, precise, and timely, addressing the traditional issues of conflicting demands between search recall and precision while balancing relevance with accuracy and timeliness(CNKI, n.d.).

The result of using 人工智能 * (道德 + 伦理) in the Subject search was further refined by selecting "Academic Journals," "CSSCI," and "CSCD," as well as the publication year from 2015 to 2024.

4.2.4.2 The EU

To search for articles from the EU on the Web of Science platform, the initial configuration was set to Search in "Web of Science Core Collection," with editions including Science Citation Index Expanded(SCI-EXPANDED) from 1945 to present and Social Sciences Citation Index(SSCI) from 1956 to present.

The "Topic" search field was utilised in Web of Science, which searches across the title, abstract, keyword plus, and author keywords(Clarivate, n.d.-a).

Then, based on the definition of search field tags(Clarivate, 2025b), the "Topic" search results using "artificial intelligence ethics" were refined by: Languages: English; Document Types: Article; Countries: All countries in the EU were selected; Publication years: 2015- 2024.

4.2.5 Final Results

Data retrieval occurred in January 2025, so all papers from 2024 are included. The paper numbers of the raw data are listed in the table4.1, indicating a total of 3218 papers, with 2339 papers from China and 879 from the EU.

[4]Beacuse the phrase "ethics of artificial intelligence" can convey the same meaning as "artificial intelligence ethics", it is preferable to avoid using the fixed term "artificial intelligence ethics" to capture all synonyms to "artificial intelligence ethics".

Table 4.1: Number of Retrieved Papers in Region and Subject

Paper Numbers	Social Science	Science	Total
China	2181	158	2339
The EU	435	444	879
Total	2616	602	3218

.

4.3 Data Preprocessing

4.3.1 Data Screening and Cleaning

To mitigate the significant imbalance between the sizes of the datasets from China(2339) and the EU(879), the papers from China were screened according to the following rules: First, papers are grouped by subject(social science; science) and year(2015 to 2024); then, 158 science papers are retained, as this number is already small compared to the 2181 social science papers. Subsets with fewer than 30 papers are also kept. Next, the top 30% papers, in order of correlation index, are selected and retained from the remaining subsets.

Then, the EU papers in the Web of Science Core Collection were bulk downloaded using Zotero (Western Michigan University Libraries, 2025). This method allows for successful access to papers that are either open-access or available through the University of Helsinki account.

Ultimately, all papers are double-checked by their titles to ensure they relate to AI ethics. The table4.2 shows the paper numbers by region and subject after screening and cleaning.

Table 4.2: Numbers of Papers in Region and Subject after Preprocessing

Paper Numbers	Social Science	Science	Total
China	705	139	844
The EU	271	327	598
Total	976	466	1442

4.3.2 Metadata Creation

Metadata was created with “title,” “year[5](from 2015 to 2024),” “subject(social science; science),” and “region(China; the EU).”

Given that the number of papers in some years is quite small, the papers were grouped into two ranges of years: 2015-2019 and 2020-2024, rather than being categorised by individual year.

[5]All papers downloaded with the original publication year of 2025 are marked as 2024 in the final metadata. This occurs because they were originally published online in 2024 but are archived in CNKI as 2025, reflecting their anticipated publication dates in journals. When the search configuration does not include the year 2025, some papers published online in 2024 will also be downloaded with a publication year labelled as 2025.

4.4 Sampling and Annotation

4.4.1 Sampling

First, eight blocks were created by combining three classifications: regions(China; the EU), subjects(social science; science), and year ranges (2015-2019; 2020-2024). Ten papers were randomly selected from each block. Some blocks contain fewer than ten papers, so all papers in those blocks were included. In the end, there are 78 sampled papers for annotation.

4.4.2 Annotation

Each sampled paper was annotated in Label Studio to indicate whether each topic from the provided list is discussed.

4.5 Modelling

4.5.1 Pilots and Model Selection

Zero-shot classification is a natural language processing task in which a machine learning model categorises data into classes it has not been previously trained to recognise (Wang et al., 2019). This method is suitable for situations where the amount of labelled data is small or no trained data is available(Yin et al., 2019). Moreover, Brown et al. (2020) illustrates that LLMs undergo training on a wide range of extensive corpora and can utilise their comprehensive language understanding to generalize across different context domains, enabling them to execute zero-shot classification effectively without requiring task-specific training. As AI ethics is already a mature area in academia and not a new field that current models are seldom trained on, the method of zero-shot classification without training a separate model is used in this research[6].

As OpenAI et al. (2024) shows, the Chatgpt model manages long-context inputs and understands text through zero-shot prompting. It exhibits human-level performance in several professional and academic domains, comprehending and generating natural language, particularly in complex and nuanced situations. GPT-4o Mini is a smaller, more affordable variant of the GPT-4o model, tailored to balance high performance with cost-effectiveness. Gemini-pro is another performance-optimised model designed for cost efficiency and low latency, with excellent reasoning capabilities and extensive multimodal functionality capabilities (Team et al., 2024).

After testing three articles across three models: GPT-4o, GPT-4o-mini, and Gemini-pro, GPT-4o was selected due to its precise performance, higher rate limits, and lowest price(Google, 2025a, 2025b; OpenAI, 2024). The comparison among the three models is shown in the table**??**, presenting their performance in terms of token limit, rate limit, price, and the accuracy of outcomes.

[6] Initial trials examined rule-based classification and Multi-Genre Natural Language Inference (MNLI)-based zero-shot text classification using the abstracts of papers, as long PDF texts are challenging to process in practice. After comparing the results with manual annotations in several pilot trials, these approaches proved unreliable in this case. One reason is that rule-based classification relies too heavily on the design of keyword lists for each topic. Additionally, the abstracts often fail to capture the complete content of the paper. These experiments were conducted in collaboration with Matia Setälä; although these two methods were ultimately not adopted, his guidance during the exploratory phase is greatly appreciated.

Table 4.3: Comparison of Models Based on Token Limit, Rate Limit, Price, and Accuracy

Model	Token Limit	Rate Limit(Tier1)	Price (per 1M Tokens)	Outcome Accuracy Order (manual check)
GPT-4o-mini	128,000	500 RPM, 200,000 TPM	$0.15 (input), $0.60 (output)	1
GPT-4o	128,000	500 RPM, 30,000 TPM	$2.5 (input), $10 (output)	2
Gemini 1.5 Pro	2,000,0000	4,000,000 TPM	$1.25 (input), $5(output)	3

4.5.2 Configurations of GPT-4o-mini

4.5.2.1 Corpus and Thematic Framework

The GPT-4o-mini model received a total of 844 papers from China and 598 papers from the EU, all in PDF format, along with the following topic lists:

Chinese Topics:

- participants: “国家机构”, “学者专家”, “公司企业”, “社会公众”, “国际组织”
- issues:“公平性”, “包容性”, “可及性”, “透明度问题”, “隐私”, “人权”, “安全性”, “严重危害”, “自主性主体性”, “不可控性”, “虚假信息”, “失业问题”, “环境危害”
- mitigation proposals: “法律法规”, “伦理道德规范”, “学术研究”, “企业责任”, “可信赖与负责任的应用”, “人工监督”, “技术解决方案”, “道德化产品设计”, “算法素养”, “公众教育”, “国际合作”

English Topics:

- participants: “Government institutions”, “Scholars and experts”, “Corporations and enterprises”, “General public”, “International organisations”
- issues: “Fairness”, “Inclusivity”, “Accessibility”, “Transparency issues”, “Privacy”, “Human rights”, “Security”, “Severe harms”, “Autonomy and agency”, “Uncontrollability”, “Misinformation”, “Unemployment” , “Environmental harms”
- mitigation proposals: “Legislation”, “Ethical and moral standards”, “Academic research”, “Corporate responsibility”, “Trustworthy and responsible use”, “Human oversight”, “Technical solutions”, “Ethical product design”, “Algorithmic literacy”, “Public education”, “International cooperation”

4.5.2.2 Prompt

Prompt: Does the following paper discuss any of these topics? Answer “Yes” or “No” for each topic.
System prompt: “role”: “system”, “content”: “You are a helpful assistant.”

4.5.2.3 Output

The output is a “Yes” or “No ” indicating whether the topic in the given list is discussed in the paper.

4.6 Evaluation of Modelling Results

The evaluation of the modelling was conducted by comparing the annotation results with the modelling results of sampled papers. First, “Yes”/“No” in both results were converted to “1”/“0”. Second, eight strata were created by combinations of three classifications: year range, region, and subject. Third, four metrics were computed for each topic within each stratum: precision, recall, F1-score, and accuracy, as they are standard in text classification assessment(Powers, 2020; Schütze et al., 2008). The F1-score was selected to assess the quality of modelling results as the harmonic mean of precision and recall. Powers (2020) noted that the F1-score is particularly suited for evaluating model performance across diverse strata, as it accounts for both false positives and false negatives. Then, a faceted box plot was generated to help identify the most problematic strata and further locate issues with the model. An F1-score of 0.65 is selected as the benchmark for good performance, consistent with the common acceptable agreement level in text annotation tasks, especially in cases where there are imbalances in topics (Goutte & Gaussier, 2005; Hripcsak & Rothschild, 2005).

4.6.1 Overall Performance

The modelling results are promising, as 72.8 % of cases meet the F1>0.65 threshold.

The faceted box plot4.1, where dots represent each topic in each stratum, shows no obvious performance imbalances between subjects. However, different regions and periods exhibit noticeably different performances. In the EU, 74.1% of the cases are considered good, whereas in China, 71.6% of the cases are classified as good, indicating that the model is more accurate in analysing English text. Additionally, the modelling results reveal a temporal decline: from 2015 to 2019, there were 76.7% good cases, compared to 2020 to 2024, which had 68.9% good cases. There are only 174 papers published from 2015 to 2019, while 1,268 papers were published from 2020 to 2024. A larger dataset potentially introduces more noise or variability and leads to inconsistent outcomes(Wallach et al., 2009).

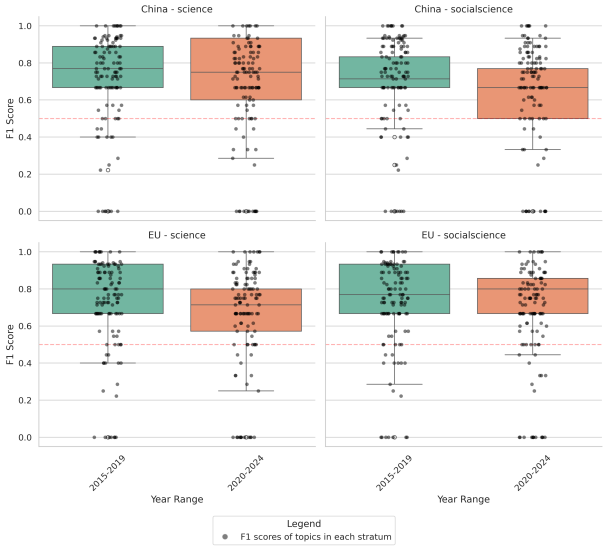


Figure 4.1: Distribution of Topics' F1 Scores in Strata by Region, Subject and Year_range

4.6.2 Main Underperformed Strata and Topics

The table4.4 shows the worst-performing strata and the key issues in each one. The most problematic stratum is the case for China's social science subjects from 2020 to 2024, with an F1 score of 0 in two topics. Additionally, science subjects are involved in two of these three strata. This may be because there are only 466 science papers, while 976 papers are from social science. Smaller datasets may not capture the full diversity of topics or patterns due to limited representation(Wallach et al., 2009). Moreover, papers in science subjects contain more specialised terminology or complexity, which makes it difficult for the model to handle(Newman et al., 2010).

Table 4.4: Worst-Performing Strata

Rank	Stratum	Avg F1	Key Issues
1	China Socialscience 2020-2024	0.38	Public education (F1=0), Algorithmic literacy (F1=0)
2	China Science 2020-2024	0.38	Environmental harms (F1=0.33), Accessibility (F1=0)
3	EU Science 2020-2024	0.40	Technical solutions (F1=0.57), Corporations and enterprises (F1=0.57)

In terms of specific topics, the table4.5 shows that the most underperforming topics are "Algorithmic literacy," "Public education," and "Environmental harm. "They all face severe imbalances in Precision-Recall when one is good (≥0.7) and the other poor (<0.5), indicating very high false positives. One possible reason is that the number of cases where "Environmental harm" is labelled as "Yes" is relatively small, so the model may struggle to identify it. Second, with high recall and low precision, the model appears overly sensitive to these topics.

Table 4.5: Underperformed Topics

Topic	Avg F1	Failure Pattern	Worst Case (F1=0)	Precision-Recall Imbalance
Algorithmic literacy	0.26	Complete failure in technical contexts	China science (all years)	Precision=0.29, Recall=1.0
Public education	0.29	High recall, near-zero precision	China socialscience 2020-2024	Precision=0.14, Recall=1.0
Environmental harms	0.15	Rare in data	6/8 cases fail	Precision=0.2, Recall=1.0

4.6.3 Speculation of Root Cause and Adjustments for Analysis

The speculated causes for the major problems are exhibited in the table4.6. “Algorithmic literacy”, “public education”, and “environmental harms” are consistently challenging across different strata. They represent situations where the model struggles to classify correctly due to complex terminology, insufficient annotation data, or inaccurate and ambiguous definitions, which can degrade the topic coherence(Newman et al., 2010). Additionally, many instances demonstrate high recall (e.g., 1.0) but low precision, suggesting that the model may over-predict specific topics, resulting in poor overall performance. In complex social contexts, such as AI ethics, ambiguous prompts can lead models to over-predict topics by capturing non-central discussions(Hovy & Spruit, 2016). The prompt used does not clearly indicate that discussing topics central to AI ethics could potentially lead to excessive false positives.

Table 4.6: Diagnostic Problems and Causes

Problem Type	Diagnostic Evidence	Examples
Inaccurate definition	Consistent failure across regions	Algorithmic literacy
Ambiguous or too broad difinition	Precision « Recall	Public education
Insufficient data	Rare in both modeling and annotation results	Environmental harms
Ambiguous prompts	High recal but low precision	Too many false postives

In summary, the model performs well in most cases. To achieve more reliable outcomes, two considerations should be noted for the next analysis of the modelling results. The illustrations on topics such as algorithmic literacy, public education, and environmental harms—the most underperforming topics—will be removed from the analysis due to their unreliability. Additionally, the results will be modified based on adjustment factors before analysis.

4.7 Preprocessing of Modelling Results

4.7.1 Calibration

Calibrating model results from machine learning, particularly for subset outcomes, involves calculating the ratios of correct predictions per topic and stratum to derive adjustment factors. The model results are

then multiplied by these ratios to enhance their precision, especially in cases of false predictions(Bella et al., 2010; Chang et al., 2009; Zadrozny & Elkan, 2002).

In the evaluation step, "Yes" or "No" in both results have been converted to "1" or "0", and eight strata have been created by combining three classifications: year_range, region, and subject. Then, the ratios of correct predictions were calculated against the total predictions for each topic in each stratum. Subsequently, the model result of each topic was multiplied by its ratio of correct prediction in the specific stratum to which it belongs. In the end, the adjusted predictions were converted back to "Yes" if the value is greater than or equal to 0.5, or "No" if the value is less than 0.5.

4.7.2 Normalization

Normalizing the data ensures that the observed trends or distributions are not simply the result of variations in the total number of papers over time, across regions, or subjects, but instead reflect actual changes in the frequencies of papers discussing specific topics. This approach is employed in natural language processing, as discussed in Wallach et al. (2009).

Process of normalisation:

- Calculate the number of papers in each stratum by region, region*year, region*subject, and region*subject*year[7]. These strata align with the research questions.
- For each topic (e.g., algorithmic literacy), sum the number of "Yes" across papers in each stratum referring to its discussion frequency.
- Calculate topic density by dividing the total number of discussions("Yes") by the number of papers in each stratum. The resulting "Topic Density, " representing the average discussion frequency of each topic per paper, serves as the final data for analysis in this research.

5 Comparison and Analysis

In this section, the focus and priority of "Issues," "Mitigation Proposals," and "Practitioners" regarding AI ethics in academic discourses are analysed by comparing the topic density across different regions(China and the EU), different periods (2015-2019 and 2020-2024), and different subjects(social science and science).

[7]Related data is listed in the Appendix.

5.1 Issues Topics on AI Ethics: Focus and Priority in China and the EU Academia from 2015 to 2024

5.1.1 Regional Comparison of Issues Topics between China and the EU Academia over the Last Ten Years

Figure5.1 compares the topic density of issues in AI Ethics between China and the EU. It shows that the autonomy&agency and privacy have emerged as common concerns for both regions. China's emphasis on "Uncontrollability" also highlights the concern about human-machine relationships, as it refers to the danger of AI going out of the control of human beings and implies the overturning of the human-machine relationship. Thus, both regions tend to discuss urgent issues related to general interest in the near future. Conversely, the less frequently discussed topics in both regions relate to issues that are not immediate threats or concern only particular groups, such as "Accessibility" and "Misinformation."

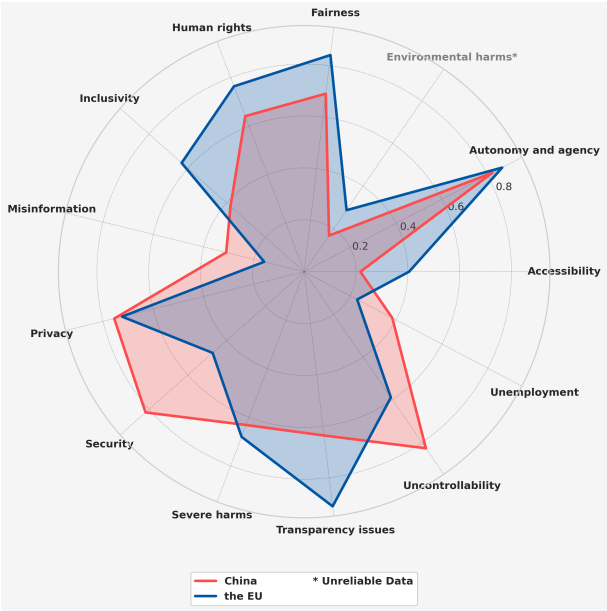


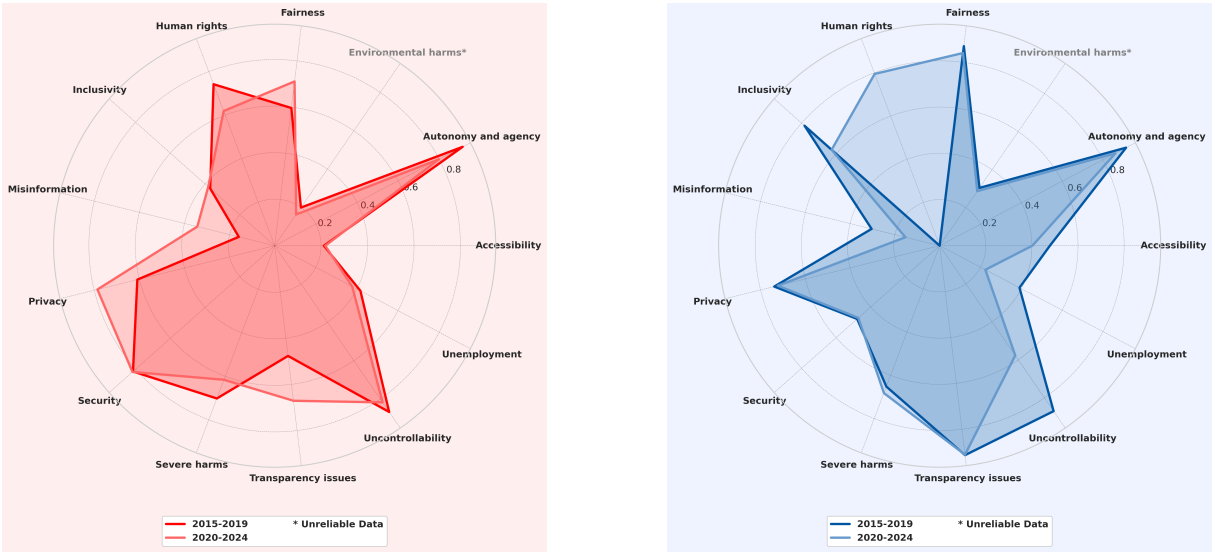
Figure 5.1: Topics Density of **Issues** in China and the EU

Regarding discrepancy, China prioritises "Security " more than the EU does. This emphasis stems from Confucianism, which strongly values societal safety and stability (Sajduk & Dziwisz, 2024). Consequently, "Security, " which refers to the reliability and resilience of AI in mitigating risks to the social system, emerges as a critical concern in China. Conversely, the EU focuses more on personal rights, highlighting issues like "Transparency issues, " "Fairness," and "Inclusivity," which have a greater topic density compared to China. This focus aligns with the human rights principles detailed in the EU Chapter of Fundamental Rights (European Union, 2012), where AI ethics are expected to support the right to access information regarding "Transparency issues" and uphold non-discrimination and equality concerning "Fairness" and "Inclusivity."

5.1.2 Temporal Changes of Issues Topics in China and the EU Academia from 2015 to 2024

Figure5.2 shows the temporal changes of AI issues in China and EU academia from 2015-2019 to 2020-2024. We can see that Chinese academia increasingly focuses on topics related to specific human rights such as "Transparency issues, " "Privacy, " and "Fairness ." This supports the fundamental national objective of creating a harmonious society that safeguards human rights while co-existing with AI(Sajduk & Dziwisz, 2024). However, due to political sensitivities, it employs the specific term "Human rights" less frequently, possibly because Western countries frequently criticise China for its human rights issues(Cath et al., 2018).

EU scholars make significant strides in "Human rights, " reflecting their consistent commitment to the EU Charter of Fundamental Rights(European Union, 2012) , that human rights should be protected in AI governance. Additionally, this implies liberal conventions in the West that emphasise the fundamental rights of individuals(Cai & Yin, 2025). On the other hand, EU academia gradually overlooks the problem of unemployment. Furthermore, concerns about uncontrollability and unpredictable AI behaviours are also fading in EU academia.



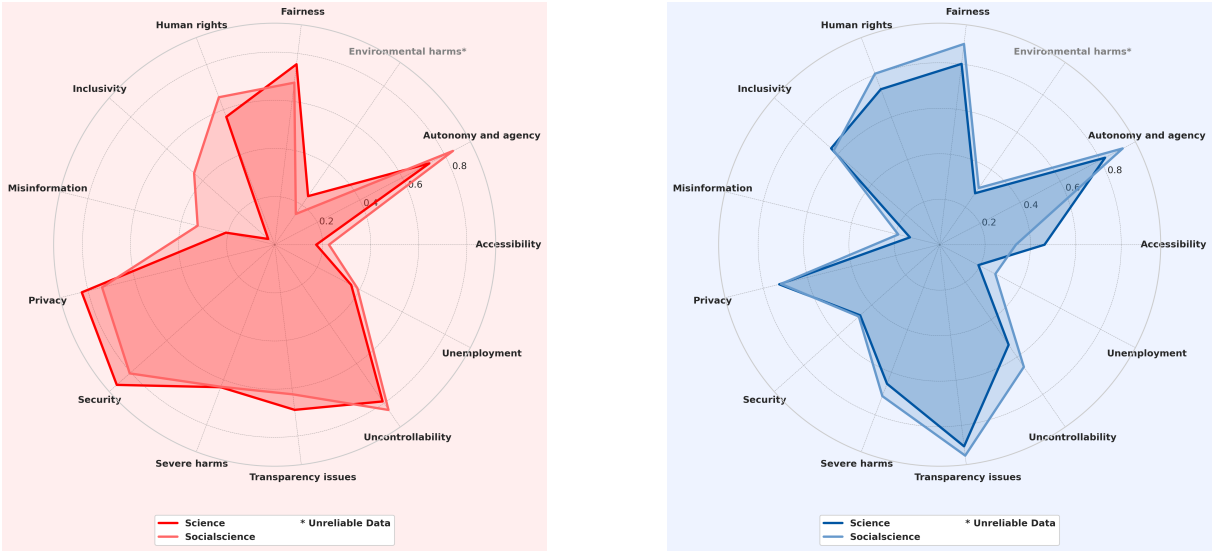
(a) China (2015-2019 vs. 2020-2024)

(b) EU (2015-2019 vs. 2020-2024)

Figure 5.2: Topic Density of **Issues** Across Two Time Periods

5.1.3 Subjects Differences of Issues Topics between Social Science and Science in China and the EU Academia from 2015 to 2024

Figure5.3 shows the overall differences in issue topics between social science scholars and science scholars in China and the EU, respectively. Scholars in both disciplines follow a similar pattern in both regions. However, Chinese social science researchers incorporate more elements from the humanities, reflecting a strengthened tradition in social science that focuses on human behaviour and societal influences. This also demonstrates the social responsibility of Chinese scholars, shaped over time by Confucianism. Historically, Confucian scholars served as advisors to rulers with the Mandate of Heaven (tianming), where leaders and intellectuals are responsible for the well-being of society(Yao, 2000).

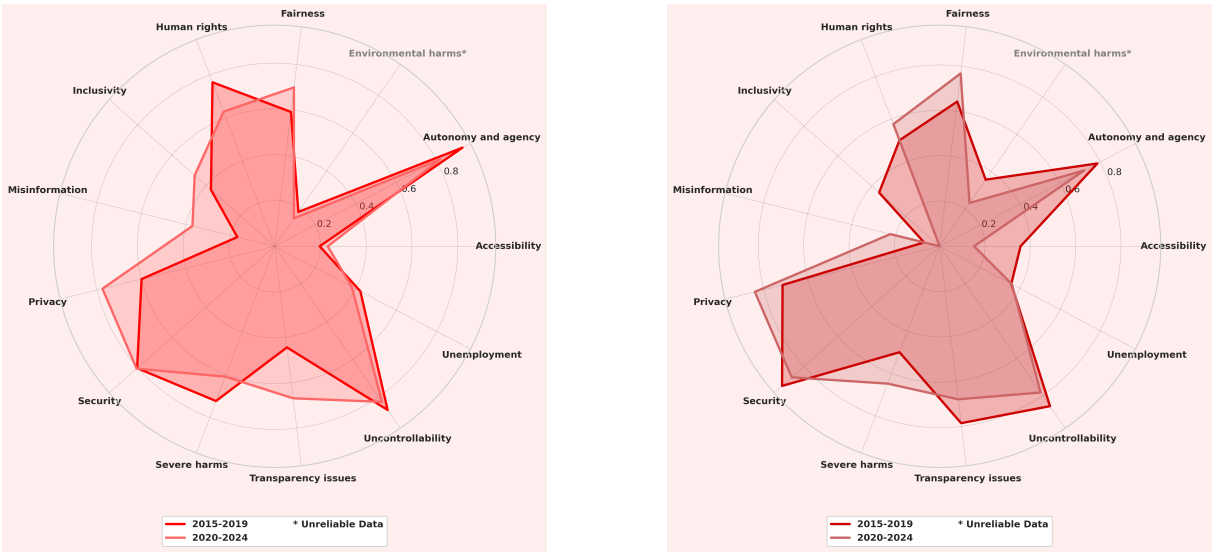


(a) China (social science vs. science)

(b) EU (social science vs. science)

Figure 5.3: Topic Density of **Issues** in Two Subjects from 2015 to 2024

Figure5.4 shows the temporal changes in issue topics discussed in China within the subjects of social science and science, respectively. Figure5.5 illustrates the same information in the EU. There are almost no differences between the two subjects in the EU. Notably, Chinese science scholars are becoming bolder and more open in discussing "Severe harms". It may be because they understand the capabilities of AI in detail and realize the potential risks earlier than those outside the technology (Zhang et al., 2021; 李修全, 2017).



(a) Social Science (2015-2019 vs. 2020-2024)

(b) Science(2015-2019 vs. 2020-2024)

Figure 5.4: Topic Density of **Issues** in **Chinese** Social Science and Science Subjects Across Two Time Periods

(a) Social Science (2015–2024)

(b) Science (2015-2019 vs. 2020-2024)

Figure 5.5: Topic Density of **Issues** in **EU** Social Science and Science Subjects Across Two Time Periods

In conclusion, the autonomy of humans and privacy have emerged as common concerns for both China and the EU. Additionally, China emphasises "Security" and "Uncontrollability" more and has shown an increasing focus on human rights-related issues in the past five years, especially among social science scholars. Chinese science scholars are discussing "Severe harms" more frequently. The EU exhibits a consistently greater emphasis on individual rights, such as "Transparency issues, " "Fairness," and "Inclusivity," but the focus on "Unemployment" and "Uncontrollability" is diminishing.

5.2 Mitigation Proposals Topics on AI Ethics: Focus and Priority in China and the EU Academia from 2015 to 2024

5.2.1 Regional Comparison of Mitigation Proposals Topics between China and the EU Academia over the Last Ten Years

Figure5.6 compares the topic density of mitigation proposals in China and the EU. Establishing moral standards and trustworthy& responsible use are the two primary proposals from academia to address AI ethics issues in both regions. They represent non-mandated approaches and serve as a means of moral compliance. Moreover, global corporations receive less attention in both regions.

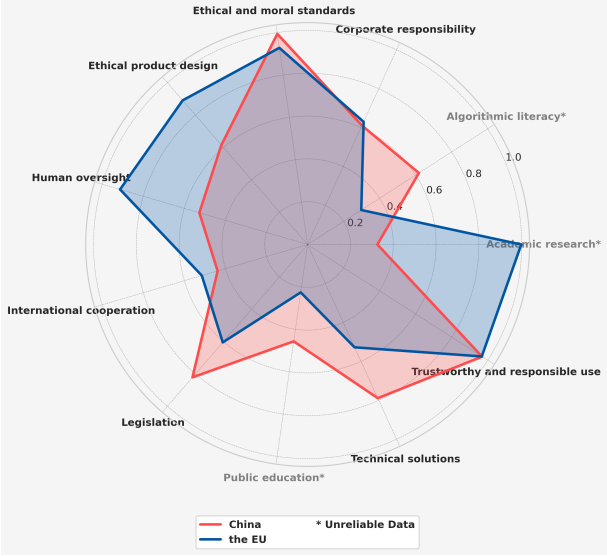
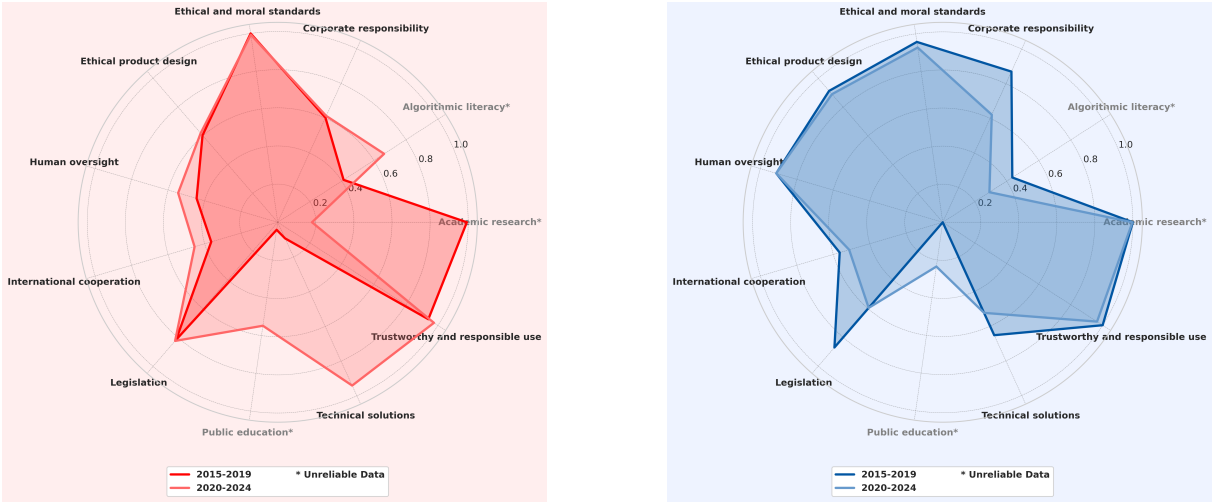


Figure 5.6: Topics Density of **Mitigation Proposals** in China and the EU

Regarding differences, Chinese academia primarily relies on legislation, highlighting the internal trust in government derived from their utopian convention. In the tenet of the utopian vision of population, there is an assumption that the government is virtuous and enjoys trust within society(Fung & Etienne, 2023), so actions taken by the government, such as legislation, are prioritised when addressing social issues. In contrast, the focus on human oversight and ethical product design in EU academia reflects their dystopian traditions, which emphasise safeguards against state abuses (Radanliev, 2025), asserting that those in power should be responsible for AI ethics and that public supervision is necessary.

5.2.2 Temporal Changes of Mitigation Proposals Topics in China and the EU Academia from 2015 to 2024

Figure5.7 depicts the temporal changes in the topics of mitigation proposals from 2015-2019 to 2020-2024 in China and EU academia, respectively. It shows a significant increase in the discussion of "Technical solutions" by Chinese scholars. This represents a competition-focused political strategy in China(Roberts et al., 2023), aimed at solving the problem of AI by developing AI. It also shows a shift towards more practical solutions in AI government, as mentioned by Sajduk and Dziwisz (2024). Meanwhile, EU scholars are shifting their focus away from actions taken by the government and private sectors, such as "Legislation" and "Corporate responsibility."



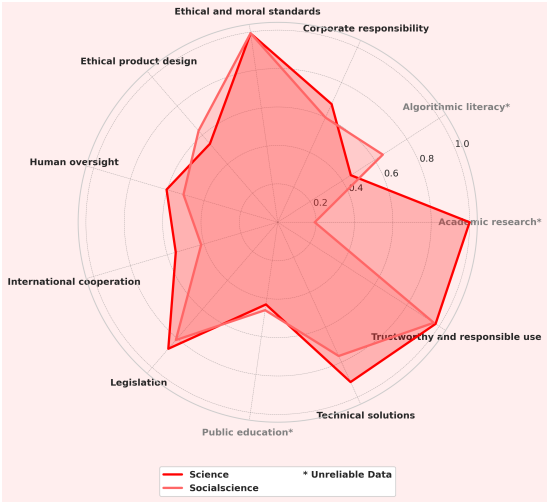
(a) China(2015-2019 vs. 2020-2024)

(b) EU(2015-2019 vs. 2020-2024)

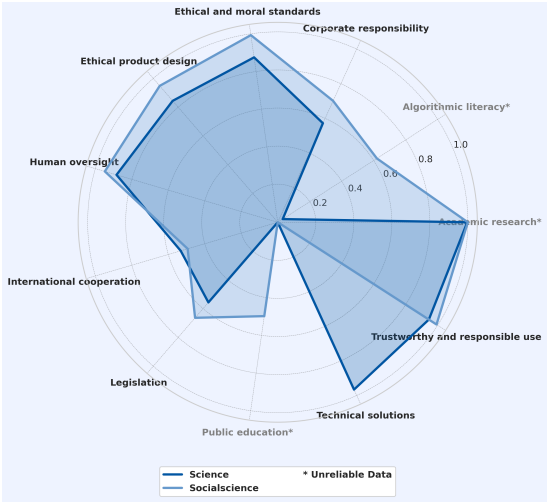
Figure 5.7: Topic Density of **Mitigation Proposals** Across Two Time Periods

5.2.3 Subjects Differences of Mitigation Proposals Topics between Social Science and Science in China and the EU Academia from 2015 to 2024

Figure5.3 shows the overall differences in mitigation proposal topics on AI ethics between social science scholars and science scholars in China and the EU, respectively. Science researchers in both regions concentrate more on “Technical solutions, ” which is reasonable considering the tradition of science scholars focusing on the practicality of technology. The primary difference is that Chinese science researchers advocate for more international cooperation than their social science counterparts. This may result from their firsthand experience with the global trend in restrictions on technology transfer(Roberts et al., 2021). China has restricted access to GPUs due to U.S. policies, which significantly hinders the infrastructure of AI(Allen, 2024). Additionally, China has limited access to global R&D ecosystems, since it prohibits some foreign applications from being in its market, and the primary AI technology is not open source. Hence, the call for breaking down barriers between countries and fostering international cooperation is a key pursuit of science scholars in China.



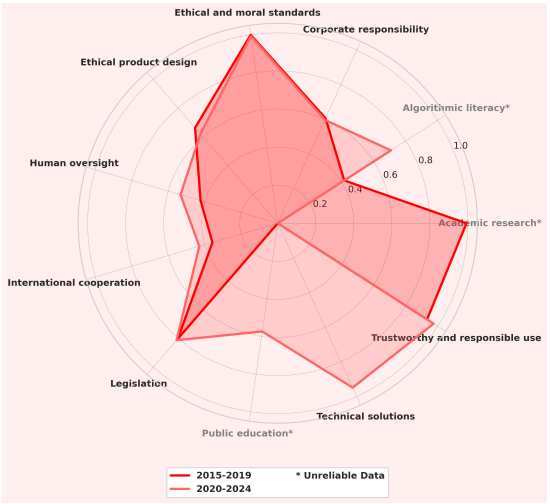
(a) China (social science vs. science)



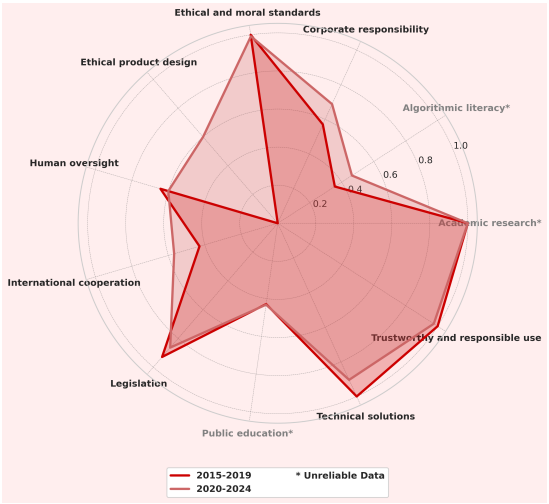
(b) EU (social science vs. science)

Figure 5.8: Topic Density of **Mitigation Proposals** in Two Subjects from 2015 to 2024

Figure5.9 shows the temporal changes in mitigation proposal topics discussed in China within the subjects of social science and science, respectively. Figure5.10 displays the same information for the EU. From 2015-2019 to 2020-2024, the most significant changes appear in China. Chinese social science researchers are increasingly acknowledging the significance of technical solutions, which align more closely with implementation solutions. Meanwhile, science researchers in China emphasise specific practices such as ethical product design to explore detailed technological possibilities. These trends reflect a shift in China that promotes more practical measures instead of focusing solely on moral compliance(贾婷 & 沈天添, 2023).

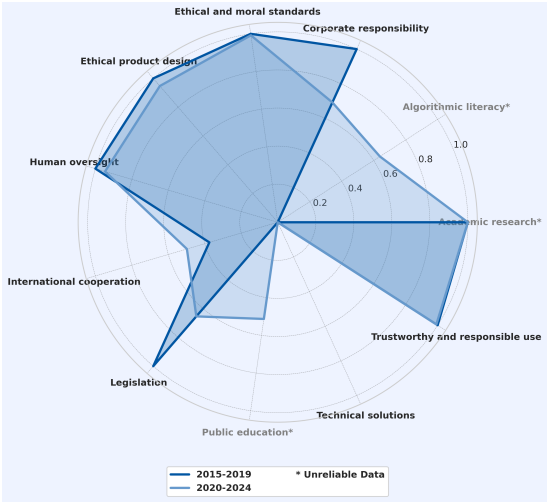


(a) Social Science (2015-2019 vs. 2020-2024)

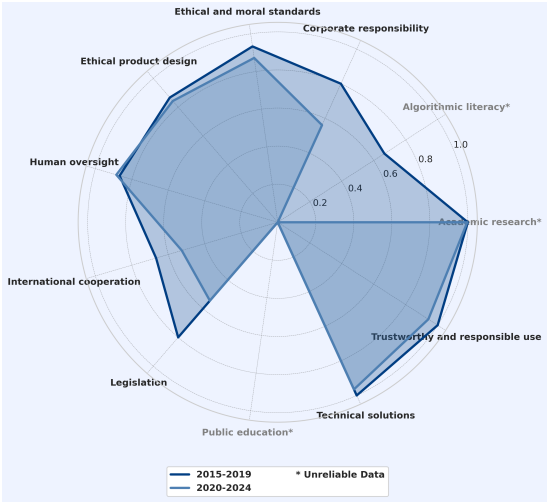


(b) Science (2015-2019 vs. 2020-2024)

Figure 5.9: Topic Density of **Mitigation Proposals** in **Chinese** Social Science and Science Subjects Across Two Time Periods



(a) Social Science (2015-2019 vs. 2020-2024)



(b) Science (2015-2019 vs. 2020-2024)

Figure 5.10: Topic Density of **Mitigation Proposals** in **EU** Social Science and Science Subjects Across Two Time Periods

In conclusion, “Ethical and moral standards” and “Trustworthy & responsible use” remain the primary proposals in both regions. Additionally, Chinese academia focuses more on “Legislation” than the EU. The EU mainly emphasises “Human oversight” and “Ethical product design”. Over the past five years, the focus of EU scholars on “Legislation” and “Corporate responsibility” has been decreasing. Chinese scholars have discussed “Technical solutions ” more, particularly social science researchers. Science researchers in China emphasise “Ethical product design” and “International cooperation” more strongly.

5.3 Practitioners in AI Ethics: Focus and Priority in China and the EU Academic Discourses from 2015 to 2024

5.3.1 Regional Comparison and Temporal Changes of Practitioners in Chinese and EU Academic Discourses over the Last Ten Years

Figure5.11 compares the overall discussion of participants in AI ethics between China and the EU. The difference is that Chinese academia focuses more on “Scholars and experts” and “Corporations and enterprises.” At the same time, the EU highlights the role of “International organisations. ” This illustrates the rising prominence of private companies and academia in the Chinese CoP(Qiao-Franco & Zhu, 2022), whereas the EU relies more on international organisations, as it is a union of countries, and European standards bodies lead the standards-setting(Roberts et al., 2023).

The results in the figure5.11 also indicate that the public is frequently discussed. However, Runcan et al. (2025) accuses both entities of insufficient public involvement. So, the high topic density of “General public” may be attributed to the fact that in academic discourses, the public is often discussed as recipients of policies or parties to be protected, rather than as participants in the decision-making process regarding AI ethics and its implementation. Nonetheless, it is a positive trend to observe more mentions of the public in the area of AI ethics.

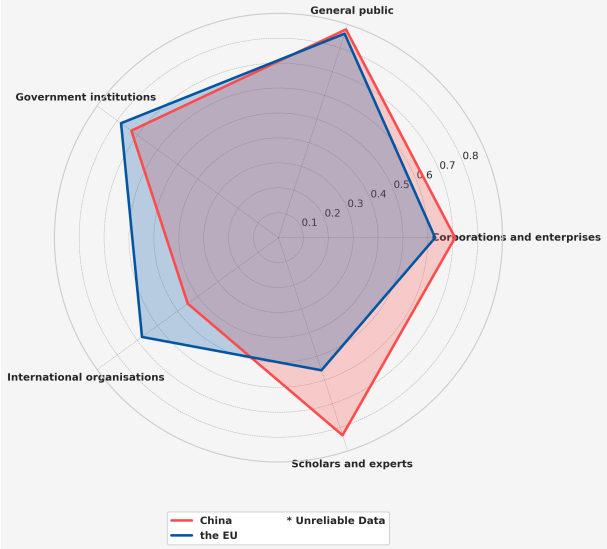
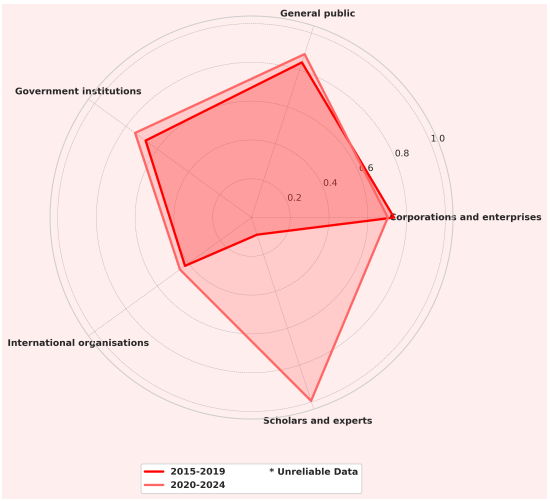
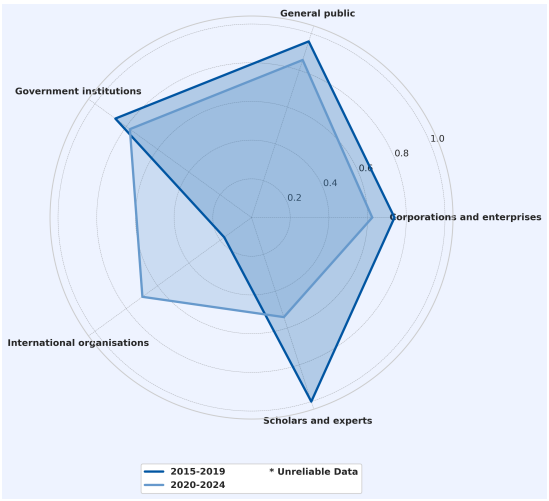


Figure 5.11: Topic Density of **Participants** in AI Ethics in China and the EU

Figure5.12 shows the temporal changes in the discussion of participants from 2015-2019 to 2020-2024 in China and the EU, respectively. It suggests that the discussion of scholars and experts has increased in China, reflecting the growing importance of academia in CoP. In contrast, the EU has emphasised international organisations more in the past five years, indicating that collaboration among member countries has become increasingly important in the EU to enhance AI ethics.



(a) China (2015-2019 vs. 2020-2024)



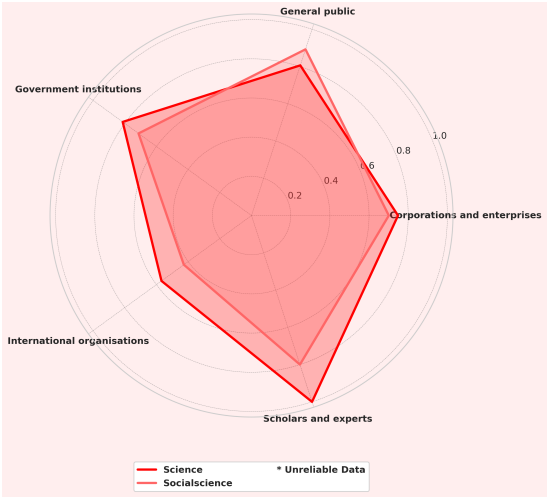
(b) EU (2015-2019 vs. 2020-2024)

Figure 5.12: Topic Density of **Participants** in AI Ethics Across Two Time Periods

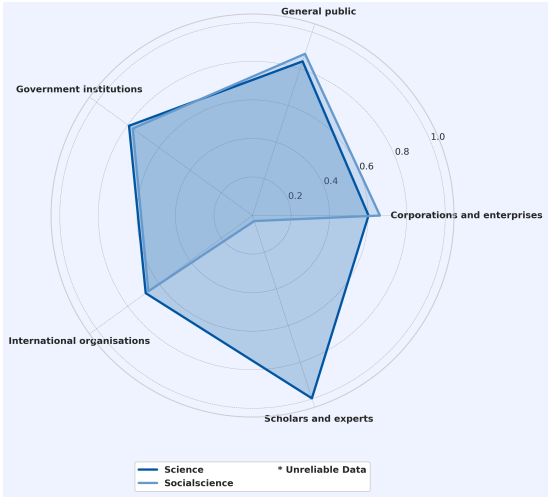
5.3.2 Subjects Differences of Practitioners between Social Science and Science in China and the EU Academic Discourses from 2015 to 2024

Figure5.13 compares the general differences in discussions about practitioners between social science scholars and science scholars in China and the EU, respectively. In China, social science researchers discuss the "General public" more than science researchers do. It reflects the nature of the humanities in

Confucianism, which calls for scholars, especially litterateurs, to care more about the living conditions of people(Yao, 2000). In the EU, science researchers show significantly greater interest in "Scholars and experts" than their social science counterparts, highlighting the predominant role of research within the science disciplines in the EU.



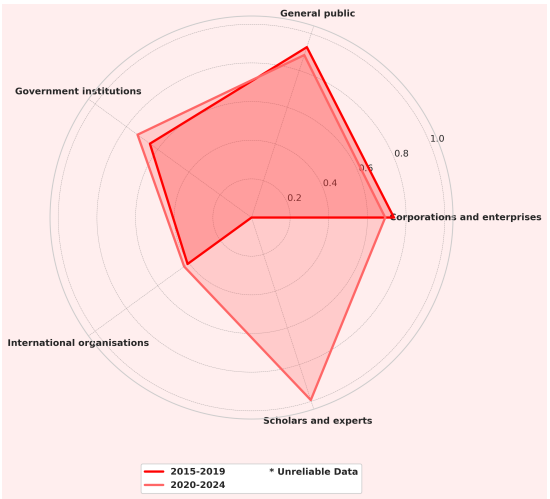
(a) China (social science vs. science)



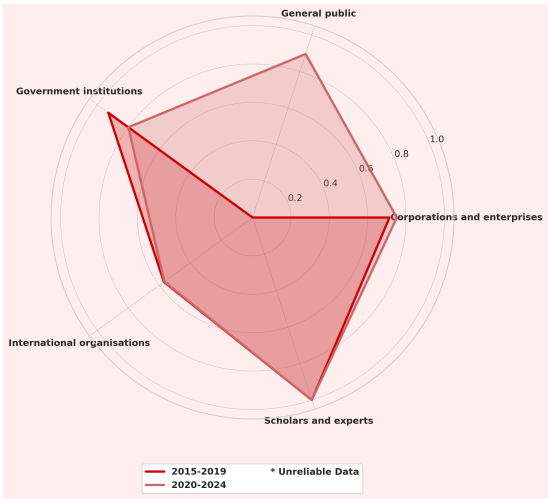
(b) EU (social science vs. science)

Figure 5.13: Topic Density of **Participants** in AI Ethics in Two Subjects

Figure5.14 shows the temporal changes in the discussion of participants in China from 2015-2019 to 2020-2024 in social science and science, respectively. Figure5.15 shows the same information for the EU. The discussion of academia is increasing in social science subjects in both regions, corresponding to the robust technical features of AI. Nonetheless, science scholars in China raise greater concerns regarding the public, mirroring the country's focus on "Harmony" to foster a more conducive human environment in the AI era (Cai & Yin, 2025). EU science researchers have emphasised the role of international organizations more in the past five years, highlighting the increasing importance of cross-border cooperation within the EU from the perspective of science scholars.

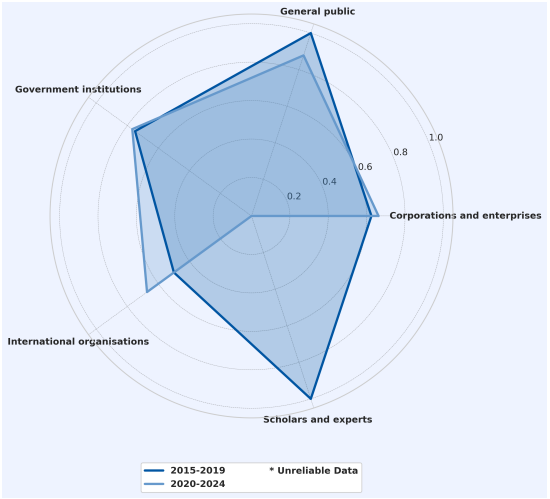


(a) Social Science (2015-2019 vs. 2020-2024)

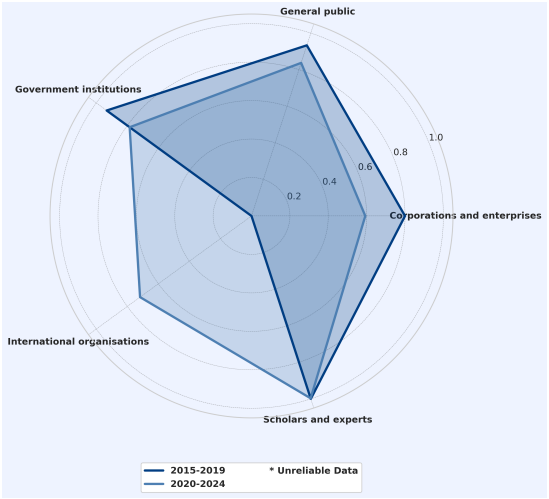


(b) Science (2015-2019 vs. 2020-2024)

Figure 5.14: Topic Density of **Participants** in AI Ethics in **Chinese** Social Science and Science Subjects Across Two Time Periods



(a) Social Science (2015-2019 vs. 2020-2024)



(b) Science (2015-2019 vs. 2020-2024)

Figure 5.15: Topic Density of **Participants** in AI Ethics in **EU** Social Science and Science Subjects Across Two Time Periods

In a word, regarding practitioners in AI ethics, Chinese academics focus more on “Scholars and experts” and “Corporations and enterprises” than those in the EU, with an increasing emphasis on “Scholars and experts” over the past five years. The EU highlights the consistently predominant role of “International organisations” from 2015 to 2024, particularly as science scholars have shown a greater emphasis on it from 2020 to 2024. In the past five years, social science researchers in both regions and science researchers in the EU have increasingly emphasised “Scholars and experts”, while Chinese science researchers express more concerns about the public.

6 Discussions and Implications

After examining the modelling results across various regions, periods, and subjects related to issues, mitigation proposals, and practitioners of AI ethics in academia, specific findings and patterns merit attention and are worthy of further discussion within the context of regional conventions.

6.1 Issues Topics in AI Ethics

The overall academic discourse in China and the EU reveals a trend of discussing AI ethics issues that concern the entire human race and are urgent both now and in the future, such as human autonomy and privacy. These issues may be more apparent with the rise of generative AI, capable of creating new content from human experience, and the future development of AGI (artificial general intelligence), designed to perform any intellectual task typically executed by humans(Everitt et al., 2018). These new developments in AI create fear among humans that AI will undermine human autonomy and make

independent decisions without human oversight. As a result, rights like human privacy can not be ensured. Meanwhile, issues that do not generate an immediate and severe impact or focus only on particular groups, like "Accessibility" and "Misinformation," are emphasised less. As Whittaker et al. (2019) has said, there is an overall trend of sidelining equality of opportunity issues in the AI era. Unemployment is no longer a key concern for AI ethics, either, and may be classified as the economic impact of AI rather than an ethical issue.

Concerning the differences between the two regions, the EU consistently focuses on human rights protections to uphold the EU Charter of Fundamental Rights, which is the constitutional legislation. It also reflects its respect for individual rights in the tradition of liberalism and concerns about the negative impact of technology stemming from its dystopian visions(Cai & Yin, 2025; Fung & Etienne, 2023; Radanliev, 2025). However, Chinese scholars are increasingly discussing fundamental human rights-related issues such as fairness, transparency, and privacy, which echo the right to equality, non-discrimination, a fair trial, and privacy in United Nations General Assembly (1948). This trend also demonstrates China's advocacy of "Harmony" in their political goals and the core tenets of Confucianism(Cai & Yin, 2025; Sajduk & Dziwisz, 2024), emphasising that people should live harmoniously with society and that public well-being should be protected as long as the entire society is thriving. Nevertheless, this protection of civic rights in China is limited by the balancing act between national stability and individual feelings. The high frequency of discussions about "Security" highlights the preferential treatment given to the safety and stability of society when faced with challenges posed by AI. It signifies a social value where the sacrifice of individuals for improved national development is deemed acceptable, rooted in China's utopian visions, which endorse the capacity and intentions of the state power(Fung & Etienne, 2023).

Regarding subject differences, scholars in the two fields exhibit similar patterns in both regions, aligning with the tradition that social science scholars are more interested in the humanities(Stenseke, 2022). However, in the past five years, Chinese science scholars have increasingly addressed severe harms, transparency, and privacy issues. One explanation is that, as they grasp the technical details, they become more aware of potential risks earlier than their social science counterparts(Zhang et al., 2021; 李修全, 2017; 赵志耘 et al., 2021). This temporal change among science scholars is not observed in the EU. This may be because the reflection of AI in Western countries began earlier than in China, due to their sceptical attitude toward technology stemming from its cyberpunk culture(Fung & Etienne, 2023).

In summary, regarding AI ethical issues, Chinese academia prioritises national interests while increasingly addressing fundamental human rights. Science scholars in China are gradually realising the severe risks posed by AI. In contrast, the EU maintains a consistent focus on human rights, which aligns with its legislative framework.

6.2 Mitigation Proposals Topics in AI Ethics

When discussing solutions to AI ethical issues, establishing moral standards and ensuring the trustworthy&responsible use of AI remain the two main proposals from academia in both regions. This suggests that academia's primary contribution lies in moral compliance and the cooperation of public and private sectors that truly control AI's application in reality (Floridi et al., 2018; Qiao-Franco & Zhu, 2022).

In terms of regional divergences, Chinese academia largely depends on legislation, indicating idealistic conventions where the public has faith in the government to wield its power virtuously for the common good (Fung & Etienne, 2023). Furthermore, while science scholars in both regions emphasise "Technical solutions," reflecting the tradition of researchers in science fields, social science scholars in China have increased their focus on "Technical solutions" as well over the past five years. This shift indicates a recent global trend towards practical solutions instead of moral guidelines (Runcan et al., 2025; 贾婷 & 沈天添, 2023). In particular, science scholars in China are calling for international cooperation, reflecting their consideration of severe global restrictions on technology transfer and the trend towards deglobalization(Sajduk & Dziwisz, 2024).

In contrast, within the legislative framework of the EU and the liberal and dystopian traditions (Cai & Yin, 2025; Fung & Etienne, 2023; Radanliev, 2025), EU academia emphasises human oversight and ethical product design to ensure public supervision and hold private companies accountable for their decisions. The EU chapter of fundamental rights requires that AI governance prioritise human rights(European Union, 2012), while the dystopian visions call for more regulations on capital and emphasise the spirit of contracts(Fung & Etienne, 2023). In the last five years, EU scholars in both subjects have discussed "Legislation" and "Corporate responsibility " less frequently. This shift might arise from the recognition of "ethics washing, " in which political and economic groups struggle to take genuine steps to enhance AI ethics when doing so contradicts their interests. This is particularly true when there are no legal mandates, even if they profess to uphold moral values standards(Hagendorff, 2020).

In summary, the mitigation proposals in academia regarding AI ethics tend to be non-binding moral norms that require cooperation from various stakeholders. Chinese scholars advocate the use of legislation to enforce the implementation of these mitigation solutions, while the EU relies more on human oversight. Both demonstrate a trend toward more practical solutions, particularly from a technical perspective.

6.3 Practitioner Topics in AI Ethics

In the academic discussion about practitioners, both regions frequently mention government institutions and the public. This reflects the government's leadership role in the field of AI, particularly in China(Qiao-Franco & Zhu, 2022). Still, this research cannot determine whether the public is involved in the decision-making process of AI ethics or remains a passive receiver.

Regarding the difference, Chinese academia emphasises "Scholars and experts" and "Corporations and enterprises" more than the EU. This difference highlights the emerging roles of private companies and academia in the Chinese CoP(Qiao-Franco & Zhu, 2022). Meanwhile, science scholars in China are becoming increasingly concerned about the public, mirroring China's Confucian traditions aimed at creating a harmonious society with responsibilities expected of scholars(Cai & Yin, 2025; Fung & Etienne, 2023). In contrast, the EU emphasises "International organisations, " and EU science researchers have also shown increased interest in it over the past five years. This focus is reasonable since the EU, as a union of countries, primarily sets its AI ethics standards through European standards bodies(Roberts et al., 2023). It also reflects the consistent emphasis on cross-border cooperation in the field of science in AI, with high involvement of technology.

In summary, the academic discussion in China aligns with the structure of the Chinese CoP in the AI sector, where the government takes the lead and the roles of industry and experts are growing. EU scholars focus on international institutions to enhance the landscape of AI ethics. This research does not clearly illustrate how the public participates in this process.

6.4 Implications for Future AI Ethics Practice

6.4.0.1 A More Inclusive and Sustainable Strategy

The main challenge identified in this research is balancing the urgency of current AI issues with long-term concerns. In addition to discussions about preserving human autonomy in the AI era and upholding general human rights, researchers should emphasise the interests of minority groups, particularly regarding accessibility. The digital divide is likely to worsen with advancements in AI technology(Lutz, 2019), making it essential to consider how minority groups, such as disabled individuals, LGBTQIA+ communities, seniors, and people from rural or low-income backgrounds, can achieve equal access to the benefits of AI. The EU has extensive experience promoting equity and inclusion, such as the regulatory practices in European Parliament and Council of the European Union (2016, 2019) from which China can learn and adapt to local needs.

Furthermore, even if misinformation and environmental harms do not have an immediate or significant impact on people's lives, stakeholders should consistently focus on these issues and maintain a long-term commitment to addressing them.

6.4.0.2 A More Practical Strategy with Balanced Relationships among Stakeholders

Examining the finalisation and implementation of AI ethics policies is vital for the near future, particularly in light of the issue of "ethical washing." In addition to moral standards, academia should engage more with AI-related institutions to expand its impact and advocate for mandated policies with legal ramifications. This trend has emerged in China with the introduction of CoP in the AI ethics field, but it requires further enhancement to manage government leadership and incorporate human oversight, particularly supervision from the public. Moreover, the issue of overly elite perspectives in the EU's tradition should also be avoided in the new communities. To enhance public involvement, AI literacy is essential. Academia can proactively promote public education to help ordinary people understand AI technology and learn how to engage with it positively.

Moreover, AI ethics should start with the innovation of AI applications by incorporating ethics into tech companies' internal R&D. This requires establishing internal AI ethics departments within companies to integrate more technical solutions into ethical product design. For example, Google's Responsible AI Team and Microsoft's Office of Responsible AI have been established to integrate AI ethics with their AI production process, offering tools and guidance through research and engineering(Google, 2025c; Microsoft, 2025).

6.4.0.3 International Cooperation and Open-Sourced AI

Despite trends of deglobalization and restrictions on technology transfer, international cooperation remains feasible when both Chinese and EU scholars seek more substantial support from international institutions and enhance communication between countries about resolving ethical issues related to AI. Furthermore, China and the EU can promote the open-source trend in AI technology to foster a culture of sharing and transparency, which also benefits the public's advancement of AI literacy through published technology details.

6.4.0.4 Interdisciplinary Research between Science and Social Science Scholars

Given the significantly greater capacity of science researchers in AI domains, they should be welcomed into AI ethics research to provide perspectives from technical viewpoints and develop more practical solutions. In addition to traditional expertise in the humanities, social science scholars can integrate insights from science researchers to create new proposals for AI ethics challenges. This form of interdisciplinary research is much better suited to AI ethics, which focuses on ethical considerations within a technological context. In this domain, researchers with a digital humanities background should assume greater responsibility by leveraging their thorough understanding of computational methods to analyse the humanities in the digital realm, merging both technological and humanistic viewpoints. The methods applied in this research reflect a classic approach to using LLMs to analyse a large-scale corpus, thereby yielding deeper and more comprehensive insights about AI ethics from academic discourses.

7 Limitations and Future Research Directions

Based on the evaluation results, the model can be improved by retraining with the data where it underperformed, including the latest data from the past five years, science-related content, and Chinese text, rather than using zero-shot classification with an existing general LLM. In addition, the model evaluation can be enhanced by adding more annotators and samples, which can also be integrated into the model's retraining and adaptations. Furthermore, simply inputting terms for topics may lead to misunderstandings about the specific content they address, as these topics have particular meanings in the context of AI ethics. Moreover, the topics examined are restricted to a specific list, limiting the discovery of new subjects. Therefore, a topic modelling method can be employed in future research to identify newly emerged topics.

Another area with potential for improvement is the prompt. It could be more accurate by explicitly requiring the model to give "Yes" only for the topic discussion that focuses solely on AI ethics themes. Therefore, further research could involve experimenting with a broader variety of prompts and inputs to address the issue of false positives.

In terms of the outcome analysis, the explanations provided by "Scholars and experts" are questionable, as the data originates from research papers. There is a risk that the model may misinterpret references to

scholars or researchers as discussions about AI ethics practitioners. Additionally, the discussion regarding the public needs deeper investigation, as their precise role as proactive practitioners or passive recipients of AI ethics implementations remains unclear in this research context.

The computational methods employed in this research to study academic discourses about AI ethics and compare large sets of Chinese and English corpora demonstrate the significant potential of LLMs in addressing humanities research questions. They enable the processing and understanding of extensive text data, facilitating more comprehensive analyses in considerably less time and resources. However, the design of the methodology and each step should be scrutinised carefully and scientifically, as a significant drawback of LLM models is that the results of each run are not reproducible.

Moreover, this research primarily focuses on a systematic overview of AI ethics discussions in academia across two regions. In the future, more in-depth studies could concentrate on specific topics of AI ethics, incorporate data from areas beyond academia, such as institutional documents and government pronouncements, and make comparisons across additional regions. With further studies, a comprehensive understanding of AI ethics and its advancement can be achieved.

8 Conclusion

This research aims to compare academic discourses on AI ethics between China and the EU from 2015 to 2024, focusing on potential differences among regions, periods, and subjects, while elucidating these divergences through the conventions of regional politics, cultures, and communities of practice. It addresses three aspects of AI ethics: issues, mitigation proposals, and practitioners; and covers two regions: China and the EU; two time periods: 2015-2019 and 2020-2024; and two subjects: social science and science. To achieve a more comprehensive and in-depth investigation, this research utilises the GPT-4o-mini model for zero-shot classification to analyse 844 Chinese papers authored by scholars in China and 598 English papers authored by scholars in the EU on AI ethics. The raw papers were retrieved from two reputable academic archives, and the search terms were modified based on previous research to create a more reliable collection of academic papers. After manually collecting and screening, 1442 academic papers on AI ethics were gathered, and a corpus was established with metadata. This latest academic dataset was then used in LLM, which effectively and efficiently analyses the content of the large-scale long-text data. Additionally, a manual annotation of 78 sampled papers was conducted to calibrate the model results and enhance the model's accuracy. Through the rigorous design of the methodology, a set of findings was identified at the end.

The scholars in both regions primarily focus on issues of human autonomy and privacy. There is a tendency to focus on the most common and urgent interests of human beings. Chinese scholars are discussing fundamental aspects of human rights more frequently, which aligns with the central topics of AI issues in the EU. This reveals a coherent core of the humanities that connects Confucianism and liberalism, as well as the requirements of the EU chapter of fundamental rights. In addition to the positive

trend of focusing on human rights, a more inclusive and sustainable strategy should be adopted by incorporating greater attention to accessibility to ensure equal opportunities for minority groups and emphasising long-term impacts, such as environmental harms, of AI development.

Regarding mitigation proposals, academic discussions in both regions emphasise establishing moral standards and the responsible&trustworthy use of AI. Chinese scholars rely more on self-regulation in political and business contexts. However, they are shifting their focus from moral compliance to technical solutions and greater involvement from technology firms in recent years. In contrast, the EU emphasises human oversight, reflecting differing levels of trust in government and perspectives on the private sector and technology between China and the EU. Nonetheless, solutions from both regions are primarily influenced by the level of cooperation from government and private companies, while public involvement remains insufficiently addressed. Hence, engaging the public is crucial for both regions to prompt the government and private sectors to fulfil their commitments. Additionally, the EU continues to operate within regulations and fosters research collaboration among nations. Meanwhile, China, particularly among science scholars, advocates for increased international collaboration to counter the trend of deglobalization. Therefore, international cooperation and calls for open-source AI represent another potential solution for improving AI ethics.

Discussions on AI ethics highlight significant parallels between the social sciences and the science fields. While social science scholars emphasise human rights, science scholars have become increasingly aware of AI's potential threats in recent years, as they have a deeper understanding of the technical aspects of AI's capabilities and risks. Thus, interdisciplinary collaboration between science and social science, particularly in digital humanities, is essential for addressing contemporary AI ethical research.

As AI develops rapidly, AI ethics will also evolve, introducing new topics while others fade in relevance. This study's reflective and innovative methodology reveals significant advancements and novel initiatives in digital humanities, indicating opportunities for exploring the ethical AI landscape in future interdisciplinary research.

9 Appendix

Table 9.1: Screened Paper Counts by Region

region	count
China	844
the EU	598

Table 9.2: Screened Paper Counts by Region and Subject

region	subject	count
China	science	139
China	socialscience	705
the EU	science	327
the EU	socialscience	271

Table 9.3: Screened Paper Counts by Region and Year_range

region	year-range	count
China	2015-2019	151
China	2020-2024	693
the EU	2015-2019	23
the EU	2020-2024	575

Table 9.4: Screened Paper Counts by Region, Subject, and Year_range

region	subject	year-range	count
China	science	2015-2019	14
China	science	2020-2024	125
China	socialscience	2015-2019	137
China	socialscience	2020-2024	568
the EU	science	2015-2019	15
the EU	science	2020-2024	312
the EU	socialscience	2015-2019	8
the EU	socialscience	2020-2024	263

References

Allen, G. C. (2024, December). China's new strategy for waging the microchip tech war. Retrieved April 13, 2025, from https://www.csis.org/analysis/chinas-new-strategy-waging-microchip-tech-war

Bakiner, O. (2023). What do academics say about artificial intelligence ethics? an overview of the scholarship. *AI and Ethics*, *3*(2), 513–525. https://doi.org/10.1007/s43681-022-00182-4

Bella, A., Ferri, C., Hernández-Orallo, J., & Ramírez-Quintana, M. J. (2010). Calibration of machine learning models. In *Handbook of research on machine learning applications and trends: Algorithms, methods, and techniques* (pp. 128–146). IGI Global. https://doi.org/10.4018/978-1-60566-766-9.ch006

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners [NeurIPS 2020]. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Cai, C., & Yin, J. (2025). Cultural and ethical foundations of ai governance divergence: A comparative analysis of china and the west. *Pol'itica Internacional*, *7*(1), 215–233. https://doi.org/10.5281/zenodo.14473453

Cath, C., Wachter, S., Mittelstadt, B., et al. (2018). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, *24*, 505–528. https://doi.org/10.1007/s11948-017-9901-7

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems* (Vol. 22). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2009/file/f92586a25bb3145facd64ab20fd554ff-Paper.pdf

Chinese Academy of Sciences. (2023). Science china [Accessed: 2023-12-01]. http://sciencechina.cn/

Clarivate. (n.d.-a). *Science citation index expanded (scie) - web of science core collection*. Retrieved October 1, 2023, from https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/science-citation-index-expanded/

Clarivate. (n.d.-b). *Social sciences citation index (ssci) - web of science core collection*. Retrieved October 1, 2023, from https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/social-sciences-citation-index/

Clarivate. (2025a). Web of science core collection: Basic search versus advanced search [Available at: https://support.clarivate.com/s/article/Web-of-Science-Core-Collection-Basic-Search-versus-Advanced-Search?language=en_US [Accessed 10 April 2025]].

Clarivate. (2025b). Web of science core collection: Search field tags [Accessed: 2025-04-22]. https://webofscience.help.clarivate.com/Content/wos-core-collection/woscc-search-field-tags.htm

Clarivate. (n.d.-a). *Advanced search - web of science core collection*. Retrieved October 1, 2023, from https://www-webofscience-com.libproxy.helsinki.fi/wos/woscc/advanced-search

Clarivate. (n.d.-b). *Web of science core collection overview*. Retrieved October 1, 2023, from https://webofscience.zendesk.com/hc/en-us/articles/26916195552401-Web-of-Science-Core-Collection-Overview

CNKI. (n.d.). *China national knowledge infrastructure*. Retrieved October 1, 2023, from https://www.cnki.net/

CNKI. (n.d.). *User guide for cnki overseas platform*. Retrieved October 1, 2023, from https://oversea.cnki.net/index/second/pdf/User%20Guide%20for%20CNKI%20Overseas%20Platform.pdf

Interim Measures for Managing Generative AI Services (2023). https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Dong, H., & Chen, J. (2024). Meta-regulation: An ideal alternative to the primary responsibility as the regulatory model of generative ai in china. *Computer Law Security Review*, *54*, 106016. https://doi.org/https://doi.org/10.1016/j.clsr.2024.106016

Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (2016, April 27). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services (2019). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L0882

Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence (Artificial Intelligence Act) (2024, June 12). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

Charter of Fundamental Rights of the European Union (2012). Retrieved October 28, 2024, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

Everitt, T., Lea, G., & Hutter, M. (2018). Agi safety literature review. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018*. https://doi.org/10.48550/arXiv.1805.01109

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3518482

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Fung, P., & Etienne, H. (2023). Confucius, cyberpunk and Mr. Science: Comparing AI ethics principles between China and the EU. *AI and Ethics*, *3*, 505–511. https://doi.org/10.1007/s43681-022-00180-6

Giarmoleo, F. V., Ferrero, I., Rocchi, M., & Pellegrini, M. M. (2024). What ethics can say on artificial intelligence: Insights from a systematic literature review. *Business and Society Review*, *129*(2), 258–292. https://doi.org/10.1111/basr.12336

Google. (2025a). Gemini API Pricing [Accessed 10 April 2025, Available in Chinese]. https://ai.google.dev/gemini-api/docs/pricing?hl=zh-cn

Google. (2025b). Gemini API Rate Limits: Tier 1 [Accessed 10 April 2025, Available in Chinese]. https://ai.google.dev/gemini-api/docs/rate-limits?hl=zh-cn#tier-1

Google. (2025c). *Responsible AI research*. Retrieved April 20, 2025, from https://research.google/teams/responsible-ai/

Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds.), *Advances in information retrieval* (pp. 345–359, Vol. 3408). Springer. https://doi.org/10.1007/978-3-540-31865-1_25

Hagendorff, T. (2020). The ethics of ai ethics: An evaluation of guidelines. *Minds & Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Han, A. G. (2008). Building a harmonious society and achieving individual harmony. *Journal of Chinese Political Science*, *13*, 143–164. https://doi.org/10.1007/s11366-008-9021-y

Hansen, C. (1992). *A daoist theory of chinese thought: A philosophical interpretation*. Oxford University Press.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine [First published: 02 April 2019]. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312. https://doi.org/10.1002/widm.1312

Hovy, D., & Spruit, S. L. (2016, August). The social impact of natural language processing. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 591–598). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-2096

Hripcsak, G., & Rothschild, A. S. (2005). Agreement, the f-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, *12*(3), 296–298. https://doi.org/10.1197/jamia.M1733

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

Lutz, C. (2019). Digital inequalities in the age of artificial intelligence and big data [First published: 26 April 2019]. *Human Behavior and Emerging Technologies*, *1*(2), 141–148. https://doi.org/10.1002/hbe2.140

Microsoft. (2025). *Responsible AI: Principles in action*. Retrieved April 20, 2025, from https://www.microsoft.com/en-us/ai/responsible-ai

Ministry of Science and Technology of China. (2019). *Governance principles for new generation artificial intelligence* (tech. rep.) (Released: 2019-06-17). MOST. https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html

Ethical Norms for New Generation Artificial Intelligence (2021). https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Civil Code of the People's Republic of China (2020). https://www.gov.cn/xinwen/2020-06/01/content_5516649.htm

Data Security Law of the People's Republic of China (2021). https://www.gov.cn/xinwen/2021-06/11/content_5616919.htm

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 100–108.

OpenAI. (2024). Gpt-4o model documentation [Accessed 10 April 2025]. https://platform.openai.com/docs/models/gpt-4o

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024, March). GPT-4 Technical Report [arXiv:2303.08774 [cs]]. https://doi.org/10.48550/arXiv.2303.08774

Powers, D. M. W. (2020). Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. https://arxiv.org/abs/2010.16061

Qiao-Franco, G., & Zhu, R. (2022). China's artificial intelligence ethics: Policy development in an emergent community of practice. *Journal of Contemporary China*, *33*(146), 189–205. https://doi.org/10.1080/10670564.2022.2153016

Radanliev, P. (2025). AI ethics: Integrating transparency, fairness, and privacy in AI development [Article number: 2463722]. *Applied Artificial Intelligence*, *39*(1). https://doi.org/10.1080/08839514.2025.2463722

Roberts, H., Cowls, J., Hine, E., Morley, J., Wang, V., Taddeo, M., & Floridi, L. (2023). Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes. *The Information Society*, *39*(2), 79–97. https://doi.org/10.1080/01972243.2022.2124565

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. In L. Floridi (Ed.), *Ethics, governance, and policies in artificial intelligence* ([Page range of chapter], Vol. 144). Springer. https://doi.org/10.1007/978-3-030-81907-1_5

Runcan, R., Hațegan, V., Toderici, O., Croitoru, G., Gavrila-Ardelean, M., Cuc, L. D., Rad, D., Costin, A., & Dughi, T. (2025). Ethical AI in social sciences research: Are we gatekeepers or revolutionaries? *Societies*, *15*(3), 62. https://doi.org/10.3390/soc15030062

Ryan, M., Antoniou, J., Brooks, L., Jiya, T., Macnish, K., & Stahl, B. (2021). Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality. *Science and Engineering Ethics*, *27*(2), 16. https://doi.org/10.1007/s11948-021-00293-x

Sajduk, B., & Dziwisz, D. (2024). Comparative analysis of ai development strategies: A study of china ' s ambitions and the eu ' s regulatory framework [Accessed via Jagiellonian University Repository]. https://doi.org/10.31175/eh4s.2014.12

Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39). Cambridge University Press Cambridge.

Sheehan, M. (2023). China ' s ai regulations and how they get made. *Horizons: Journal of International Relations and Sustainable Development*, (24), pp. 108–125. Retrieved April 21, 2025, from https://www.jstor.org/stable/48761167

Personal Information Protection Law of the People's Republic of China (2021). https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm

Stanford Institute for Human-Centered Artificial Intelligence (HAI). (2025). *Ai index report 2025* (Accessed: 2025-04-15). Stanford University. https://hai.stanford.edu/ai-index/2025-ai-index-report

State Council of China. (2016). Cybersecurity law of the people's republic of china [Published: 2016-11-07, Effective: 2017-06-01]. https://www.gov.cn/xinwen/2016-11/07/content_5129723.htm

Stenseke, J. (2022). Interdisciplinary confusion and resolution in the context of moral machines. *Science and Engineering Ethics*, *28*(3), 24. https://doi.org/10.1007/s11948-022-00378-1

Team, G., Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., Silver, D., Johnson, M., Antonoglou, I., Schrittwieser, J., Glaese, A., Chen, J., Pitler, E., Lillicrap, T., Lazaridou, A., … Vinyals, O. (2024, June). Gemini: A Family of Highly Capable Multimodal Models [arXiv:2312.11805 [cs]]. https://doi.org/10.48550/arXiv.2312.11805

Tran, B. X., Vu, G. T., Ha, G. H., Vuong, Q.-H., Ho, M.-T., Vuong, T.-T., La, V.-P., Ho, M.-T., Nghiem, K.-C. P., Nguyen, H. L. T., Latkin, C. A., Tam, W. W. S., Cheung, N.-M., Nguyen, H.-K. T., Ho, C. S. H., & Ho, R. C. M. (2019). Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. *Journal of Clinical Medicine*, *8*(3), 360. https://doi.org/10.3390/jcm8030360

Triandis, H. C. (1993). [review of the book *Cultures and Organizations: Software of the Mind*, by g. hofstede]. *Administrative Science Quarterly*, *38*(1), 132–134. https://doi.org/10.2307/2393257

United Nations General Assembly. (1948). Universal Declaration of Human Rights [Adopted 10 December 1948, Resolution 217 A (III)]. Retrieved April 18, 2025, from https://www.un.org/en/universal-declaration-human-rights/

University, N. (n.d.). *Chinese social sciences citation index (cssci)*. Retrieved October 1, 2023, from https://cssrac.nju.edu.cn/cpzx/zwshkxywsy/sjkjj/20191231/i63997.html

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *Proceedings of the 26th Annual International Conference on Machine Learning*, 1105–1112. https://doi.org/10.1145/1553374.1553515

Wang, W., Zheng, V. W., Yu, H., & Miao, C. (2019). A Survey of Zero-Shot Learning: Settings, Methods, and Applications. *ACM Transactions on Intelligent Systems and Technology*, *10*(2), 1–37. https://doi.org/10.1145/3293318

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., … Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv*, *abs/2112.04359*. https://doi.org/10.48550/arXiv.2112.04359

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Hendricks, L. A., Rimell, L., Isaac, W., … Gabriel, I. (2022). Taxonomy of risks posed by language models. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 214–229. https://doi.org/10.1145/3531146.3533088

Western Michigan University Libraries. (2025). *Zotero: Saving full-text pdfs automatically*. Retrieved March 30, 2025, from https://libguides.wmich.edu/zotero/fulltext

Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, E., Mills, M., Morris, M. R., Rankin, Y., Rogers, E., Salas, M., & West, S. M. (2019). Disability, bias, and ai. Retrieved April 13, 2025, from https://ainowinstitute.org/publication/disabilitybiasai-2019

Yao, X. (2000). *An introduction to confucianism*. Cambridge University Press.

Yin, W., Hay, J., & Roth, D. (2019, August). Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach [arXiv:1909.00161 [cs]]. https://doi.org/10.48550/arXiv.1909.00161

Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 694–699. https://doi.org/10.1145/775047.775151

Zhang, B., Anderljung, M., Kahn, L., Dreksler, N., Horowitz, M. C., & Dafoe, A. (2021). Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers. *Journal of Artificial Intelligence Research*, *71*. https://doi.org/10.1613/jair.1.12895

Zhu, J. (2024). AI ethics with Chinese characteristics? Concerns and preferred solutions in Chinese academia. *AI & SOCIETY*, *39*(3), 1261–1274. https://doi.org/10.1007/s00146-022-01578-w

吴汉东. (2017). 人工智能时代的制度安排与法律规制. *法律科学 (西北政法大学学报)*, *35*(5), 128–136.

李修全. (2017). 人工智能应用中的安全, 隐私和伦理挑战及应对思考. *科技导报*, *35*(15), 11–12. Retrieved October 31, 2024, from http://www.kjdb.org/CN/article/downloadArticleFile.do?attachType=PDF&id=14456

段伟文. (2017). 人工智能时代的价值审度与伦理调适. *中国人民大学学报*, *31*(6), 98–108.

贾婷, 陈., & 沈天添. (2023). 基于 LDA 模型的人工智能伦理准则体系研究. *同济大学学报（自然科学版）*, *51*(5), 652–659.

赵志耘, 徐峰, 高芳, 李芳, 侯慧敏, & 李梦薇. (2021). 关于人工智能伦理风险的若干认识. *中国软科学*, *6*, 1–12.

马长山. (2018). 人工智能的社会风险及其法律规制. *法律科学 (西北政法大学学报)*, *36*(6), 47–55.