

Research on the Ethical Constraint Mechanism for Information Value Development in the Age of Artificial Intelligence: Postprint

Authors: Yang Danxiu

Date: 2023-10-08T00:00:00+00:00

Abstract

As artificial intelligence continues its development, the risks in data ethics and algorithmic ethics induced by the technology have become topics of intense contemporary discussion. Data ethics risks primarily encompass issues of personal privacy security, information security, ambiguity in the boundaries of data rights, and the intelligent divide; algorithmic ethics risks mainly include algorithmic bias, algorithmic discrimination, and the negative influence and information cocoon risks inherent in algorithmic recommendation. These ethical risks urgently require resolution through mature constraint mechanisms, which can be primarily examined from three dimensions: the promulgation of laws and regulations centered on the definition of data rights; the construction of ethical mechanisms that combine principal responsibility, industry self-regulation, and citizens' network literacy; and technological innovation in algorithmic programs.

Full Text

Preamble

Research on Ethical Constraint Mechanisms for Information Value Development in the Artificial Intelligence Era

(School of Media and Communication, Yunnan University of Finance and Economics, Kunming, Yunnan 650221)

Abstract: As artificial intelligence continues to develop, the data ethics and algorithmic ethics risks triggered by its technologies have become hotly debated topics. Data ethics risks primarily include personal privacy security, information security, blurred boundaries of data rights, and intelligent divides. Algorithmic ethics risks mainly encompass algorithmic bias, algorithmic discrimination, and the negative induction and information cocoon risks implied by algorithmic

recommendation. These ethical risks urgently require mature constraint mechanisms for resolution, which can be addressed through three dimensions: the promulgation of laws and regulations centered on the definition of data rights; the construction of ethical mechanisms combining subject responsibility, industry self-discipline, and citizens' network literacy; and technological innovation in algorithmic programs.

Keywords: artificial intelligence; data ethics; algorithmic ethics; constraint mechanisms

CLC Number: TP392

Document Code: A

Article ID: 1671-0134(2021)07-068-05

DOI: 10.19483/j.cnki.11-4653/n.2021.07.019

Citation Format: Yang Danxiu. Research on Ethical Constraint Mechanisms for Information Value Development in the Artificial Intelligence Era [J]. China Media Technology, 2021(07): 68-72.

1. The Generative Logic of AI Ethical Risks in Information Value Development

In 1956, at the Dartmouth Symposium held in Dartmouth, USA, a group of young scientists led by Minsky, McCarthy, and Newell gathered together to discuss topics related to machine simulation of intelligent technologies [1]. This conference marked the birth of artificial intelligence as an emerging science and technology. While AI data and algorithmic analysis provides technical support for human production and life, it has also induced numerous data ethics and algorithmic ethics problems that may affect social stability and fairness. For instance, some criminals may illegally steal others' private data to carry out criminal activities, such as illegal human trafficking on the dark web, while unconstrained algorithmic systems may generate algorithmic bias against specific groups.

For example, on February 17, 2020, Xinhua News Agency disclosed that during the Spring Festival, criminals used big data methods to leak and repost large amounts of personal information of people who returned to Wuhan, threatening the information security of numerous individuals.

1.1 The Generative Logic of Data Ethical Risks

Data ethics risks involve ethical issues caused by large-scale data collection and development, primarily including personal privacy and information security problems, damage to data rights under data hegemony, and intelligent divides caused by information monopolies.

1.1.1 Generative Factors First, the complexity of data sources. Data records of network users' browsing and usage traces carry subjective biases of both data consumers and collectors to a certain extent. The volume of training data in algorithmic programs also affects the value orientation of data output. For example, in crime statistics released by the FBI in 2016, the crime rate among Black people reached 37% of the total 408,871 crimes, which was significantly higher than that of white people. However, the underlying reasons are controversial, as violent tendencies resulting from long-term discrimination remain debatable, and group characteristics cannot represent the entire Black population.

Second, the blurring of data rights boundaries. Platform-accumulated user data includes both data from users' active browsing traces and passive data generated when users consume information with certain attached conditions. These conditions are often designed to circumvent users' informed consent, reflecting a lack of ethical norms and constraints between platforms and application subjects. For instance, in July 2018, Dianping used web crawler technology to capture large amounts of personal information from Xiaohongshu users, plagiarizing and reposting original notes from the platform. Such illegal theft of user data by internet enterprises not only infringes on users' personal rights but also leads to imbalances in the data application market, losing space for benign development.

Third, the imbalance in data application markets. In terms of AI big data development and application in the domestic market, big data development remains in a relatively chaotic stage. There exists unhealthy competition in data transparency and openness, and no reasonable and effective market constraint mechanism has emerged between the concealment of data mining and de-privatization.

1.1.2 Generative Mechanism Data originates from individuals who are deeply embedded in cultural contexts, and cultural consciousness itself carries subjectivity. Once structural prejudice against a minority group exists in society, these biases will inevitably be reflected in the data. Regarding data sources, the boundary definition of data groups presents another problem during platform development and application. Users cede partial rights to obtain information services provided by platforms, while some platforms increasingly engage in data abuse lacking ethical norms. The spontaneity and profit-seeking nature of the data market have sensed this opportunity, with more platforms using unfair competition methods to scrape user data, causing imbalances in the data application market [Figure 1: see original paper].

[Figure 1: see original paper] Schematic Diagram of Data Ethics Risk Generation Mechanism

1.1.3 Classification of Data Ethics Risks First, privacy security issues. In discussions of data ethics risks, privacy issues constitute the core. The development of AI technology relies on big data collection, which may involve

secondary use of personal information, illegal collection, and information leakage and diffusion. In September 2018, Xinhua News Agency reported that telephone fraud gangs using numbers starting with 9521 generally collected personal details through various internet channels, with harassed users reaching 4.1 million and threatening information security.

Second, information security issues. With internet popularization and AI technology development, personal information faces constant leakage risks. In July 2020, the Beijing Internet Court ruled in its first-instance judgment that both WeChat Reading and Douyin Apps infringed on users' personal information. When big data integrates and captures personal information, it can easily leak relevant personal data, leaving people seemingly living in a Truman Show where every word and action on websites and social media risks being monitored.

Third, user data rights damage issues. According to the connotation of data rights, data collection and use require users' informed consent. Unrestrained data sharing only leads to damage to data subjects' rights and chaos in data order. Data rights include the right to informed consent, privacy, deletion, and data portability as specific manifestations of basic human rights in the data domain. In April 2016, hackers attacked Turkey's national information security database, leaking information of 50 million citizens, including President Erdogan, online, while the information security department failed to properly protect sensitive personal data and access permissions.

Finally, data monopoly and digital divide issues. For data developers, monopolizing data leads to uneven rights distribution and data infringement. Foreign giants like Facebook and Google, and domestic giants like Tencent and Alibaba, possess extensive information sources and technological advantages, monopolizing most data resources in the application market. For data consumers, compared with those without internet access, users who access the internet more frequently can receive more effective information provided by network operators. According to the 2019 Global Internet Digital Report, as of December 2018, global internet users had reached 4.1 billion, while approximately 48% of the population remained unconnected, meaning people do not enjoy fair data distribution rights, and the knowledge gap between internet users and non-users will increasingly widen.

1.2 The Generative Logic of Algorithmic Ethical Risks

Algorithmic ethics risks refer to unforeseen or potential adverse consequences generated when developing data through specific instructions and programs [3]. The main causes include algorithmic uncontrollability, subjective value preferences in algorithm development and application, and imperfect ethical constraint mechanisms. Primary risks involve negative induction from algorithmic recommendations, information cocoon problems, algorithmic bias, and algorithmic discrimination.

1.2.1 Generative Factors First, algorithmic uncontrollability. Machine learning can autonomously write code through appropriate algorithms and sufficient training data, using data to generate new patterns and knowledge. Based on big data systems, decision rules can be effectively modified, extended, and adjusted. However, different data varies in scale, quality, and source, affecting algorithmic program complexity, making some algorithm-derived risks unpredictable. For example, in 2016, Facebook’s algorithm identified the 1972 anti-war work “Napalm Girl” as inappropriate content, leading to account bans and enormous public pressure. The platform struggled to balance autonomous algorithmic adjustment with ethical risk management.

Second, subjective value preferences in algorithm development and application. Algorithmic decisions often use historical data to predict future trends, where algorithmic model design and data input affect prediction results. Developers may embed their subjective will into algorithmic systems. For users, after algorithm development, they can independently set parameters according to their objective needs when using smart devices. The book “The Beauty of Algorithms” mentions an overfitting example where detailed personal opinion-related parameters caused missed sampling data points and failed predictions. Similarly, algorithmic programs overly influenced by designers’ and users’ subjective value preferences can cause unnecessary risks.

Third, imperfect ethical constraint mechanisms. Algorithms can autonomously adjust programs and parameters, adding uncertainty. Meanwhile, people have yet to further explore and research direct responsibility subjects and controllable mechanisms for algorithmic programs. As algorithm development and application platform subjects, lacking perfect ethical constraints will affect the benign cycle of the entire internet ecosystem and induce more algorithmic ethics risks. For example, in April 2018, ByteDance was penalized and permanently shut down its product “Neihan Duanzi” due to vulgar content and value orientation problems, and user comments on its subsidiary platform “Douyin” were also cleared.

1.2.2 Generative Mechanism During algorithm development and design, AI technologies like machine learning can autonomously write code through appropriate algorithms and training data to execute common information service instructions. However, lacking deep analysis and value recognition of human society’s cultural connotations, they often exceed algorithm and platform designers’ ethical regulations and derive ethical risks. During application, designers’ subjective will and users’ autonomous modification of service parameters further affect errors in collected data and new algorithmic program norms. Additionally, platform algorithmic mechanisms lacking comprehensive ethical constraints in value orientation and content orientation contribute to risk generation [Figure 2: see original paper].

[Figure 2: see original paper] Schematic Diagram of Algorithmic Ethics Risk Generation Mechanism

1.2.3 Classification of Algorithmic Ethics Risks First, algorithmic bias issues. Algorithmic bias refers to negative cognition and attitudes toward people, events, and things established during algorithmic decision-making [4]. *Nature* magazine calls this form, where decision processes cannot be understood before a “black box” and only results are comprehensible, “bias in, bias out.” For instance, in 2015, Google’s image recognition software erroneously labeled Black people as “apes” or “chimpanzees,” triggering a series of racial discrimination controversies.

Second, algorithmic discrimination issues. Algorithmic discrimination most commonly manifests as big data-based price discrimination, where operators adopt different pricing standards for different consumers when selling the same product or service to capture more consumer surplus. For example, in December 2020, multiple consumers reported that Tmall Supermarket displayed higher prices for Tmall 88 VIP members than for regular users.

Finally, negative induction and information cocoon problems from algorithmic recommendations. Precision algorithmic recommendations based on big data systems may pose negative induction risks. For instance, in March 2018, Cambridge Analytica collaborated with then-US President Trump during the presidential election, illegally obtaining data from 50 million Facebook users and implementing targeted election information pushes to interfere with elections in some countries, suspected of negatively inducing user voting behavior. The information cocoon is a phenomenon arising from intelligent communication. In general communication contexts, it mainly refers to how platform technical logic leads to single-sided and one-sided information pushes for users, gradually causing user self-enclosure and information narrowing over time. While algorithmic precision recommendations achieve accurate demand matching based on user profiles, they greatly accelerate information asymmetry between platforms and users and among users themselves, easily leading to spiritual control of consumers by single information.

2. Data Ethical Constraints in Information Value Development

Data ethics issues include data mining, storage, and utilization. Due to the complexity of data sources, imbalance in data application markets, and blurred boundaries of data rights, data ethics problems such as information security, data hegemony, and information divides easily arise, necessitating constraints on data sources, data rights boundaries, and data application markets.

2.1 Constraints on Data Sources

Data has uncertain, dynamic, and incomplete characteristics, yet training data in algorithmic program design is relatively simple and statically distributed in sampling information, making collected data inherently limited. Additionally,

data records of network users' browsing and usage traces carry subjective biases of data consumers and collectors. For example, machine learning algorithms may not consider that attackers could intentionally feed malicious data to machine learning systems, causing trained models to deviate from reality. Since AI involves many related technologies, such security problems will be widely prevalent.

2.2 Constraints on Data Rights Boundaries

The definition of data power boundaries changes according to specific contexts, and new rights definition issues emerge during these changes, complicating data rights attribution. Most platform-accumulated user data can be divided into two categories: active data voluntarily retained by users, such as information queries and subscriptions through intelligent search engines, and passive data generated when users consume information with attached conditions, such as when mobile Apps request access to location and contact lists. These conditions are often designed to circumvent users' informed consent rights and lower data acquisition thresholds, reflecting a lack of ethical norms and constraints between platforms and application subjects.

2.3 Constraints on Data Application Markets

AI big data development and application in the domestic market remain in a relatively chaotic stage. There exists unhealthy competition in data transparency and openness, and no reasonable and effective market constraint mechanism has emerged between the concealment of data mining and de-privatization [5]. For example, facial recognition technology in computer vision is widely applied, and improper handling of collected user data may harm user interests. On the other hand, users' browsing traces on the internet become new data information for 沉淀 and collection, creating risks of personal privacy data leakage during this process.

3. Algorithmic Ethical Constraints in Information Value Development

Algorithmic ethics involves how AI, machine learning, and robots interpret algorithms and whether they can provide correct decisions when people face moral dilemmas. Constraints on algorithms mainly address uncontrollability, subjective value preferences in development and application, and imperfect ethical constraint mechanisms.

3.1 Constraints on Algorithmic Uncontrollability

Machine learning can autonomously write code through appropriate algorithms and sufficient training data, using data to generate new patterns and knowledge.

Based on big data systems, decision rules can be effectively modified, extended, and adjusted [6]. However, different data varies in scale, quality, and source, affecting algorithmic program complexity and making some algorithm-derived risks unpredictable.

3.2 Constraints on Subjective Value Preferences in Algorithm Development and Application

Algorithmic decisions often use historical data to predict future trends, where algorithmic model design and data input affect prediction results, and algorithms essentially represent mathematical or code-based opinion expression [7]. For designers, factors like algorithmic design, data screening, and application standards imply developers' subjective attitudes, and they may embed their subjective will into algorithmic systems. For users, after algorithm development, they can independently set parameters according to their objective needs when using smart devices.

3.3 Construction and Constraints of Ethical Mechanisms

Algorithms can autonomously adjust programs and parameters, adding uncertainty to algorithmic programs. Meanwhile, people have yet to further explore and research direct responsibility subjects and controllable mechanisms for algorithmic programs. Scientists like Hawking and Schmidt have warned that strong AI or superintelligence may threaten human survival. From Asimov's Three Laws of Robotics to the 23 AI principles proposed at the 2017 Asilomar Conference, discussions on potential ethical safety issues of AI have been continuous. However, few theoretical mechanisms have specifically addressed subject responsibility mechanisms combining current algorithmic program development realities. Therefore, establishing a scientific and complete ethical constraint mechanism is particularly important in facing new risks and challenges from data ethics and algorithmic ethics.

4. The Value of AI Ethical Constraints in Information Value Development

The value of AI ethical constraints in information value development can be divided into three aspects: protecting personal data security and respecting human rights; better leveraging the positive role of big data technology and algorithms while grasping the boundaries of social information security; and guiding the reshaping of information value ethics concepts to promote human comprehensive development.

4.1 Protecting Personal Data Security and Respecting Human Rights

Everything will be recorded and analyzed, and negative ethical concepts amplify the impact of data and algorithms on personal privacy security. Algorithm goals are designed by humans, and algorithms themselves do not possess the subjective will to destroy humanity. Our focus should be on how to ensure data and algorithmic program development stays in the hands of ethically normative designers to better serve humanity. “From an individual perspective, information security in the AI era encompasses freedom, dignity, and ethical values” [8]. In the intelligent era, users’ personal information data is closely related to freedom and rights. Since personal rights like freedom and dignity are recognized ethical values, and whether personal privacy can be protected implies whether these rights are respected, protecting personal information security in the intelligent era has important ethical value at the individual level.

4.2 Better Leveraging the Positive Role of Big Data Technology and Algorithms, and Grasping the Boundaries of Social Information Security

These ethical risks accompany the technical dividends brought by AI. From a social perspective, AI information security has positive significance for achieving ethical values like civilization and harmony. Civilization and harmony are essential for social existence and development and are effective value goals justified by morality. For instance, modern social organizations with massive data easily become targets for hackers, and data leakage risks can adversely affect corporate reputation and social stability, making information security boundaries more clearly defined. Conversely, information security deficiencies harm social civilization and harmony.

4.3 Guiding the Reshaping of Information Value Ethics and Promoting Human Comprehensive Development

“Ethical concepts from the small data era lack universal applicability for constraining ethical risks in today’ s AI development, inspiring us to establish new ethical concepts of fairness, justice, openness, diversity, and mutual sharing” [9]. On one hand, data and algorithm developers should emphasize compliance with relevant laws, regulations, and industry norms being perfected. On the other hand, data subjects should enhance their data capabilities and network literacy, pursue new information freedom through people-oriented data rights, and achieve mutual benefit and win-win outcomes. Through effective data ethics and algorithmic ethics constraint mechanisms, users can be guided to access and understand diverse information choices, preventing information bias from affecting human comprehensive development.

5. Challenges of Intelligent Ethical Risks and Constraint Mechanisms in Information Value Development

Challenges of intelligent ethical risks in information value development include: first, the definition of personal privacy protection; second, the design of algorithmic recommendation mechanisms; and third, the complexity of data information selection. To address these issues, we must first strengthen protection of personal privacy information; second, construct internal and external algorithmic constraint mechanisms; and finally, improve AI technological innovation capabilities.

5.1 Challenges of Intelligent Ethical Risks in Information Value Development

First, the definition of personal privacy protection. Different countries and ethnic groups have different definitions of privacy issues. UNESCO's "Preliminary Draft Report on Robot Ethics" passed in 2016 stated that damage caused by intelligent technology and its derivative products to humans can be adjusted through product liability systems, treating intelligent products as general technological products. However, these opinions only address weak AI products and lack universal applicability for future intelligent technology products with independent consciousness.

Second, the design of algorithmic recommendation mechanisms. Different users have different information needs, and algorithmic recommendations can save time and selection costs for users with intense life and work rhythms but also trouble users needing diverse information. From a technological innovation perspective, many technical risks remain in technology development and application stages, such as personal privacy protection and data bias.

Third, the complexity of data information selection. Since user-generated data information is complex, some useless data is retained while important data is deleted, creating challenges for data cleaning. Data developers' subject responsibility constraint mechanisms need improvement, lacking self-discipline norms for inter-institutional data development, while consumers have weak rights awareness.

5.2 Ethical Constraint Mechanisms for Information Value Development

First, improve laws, regulations, and AI supervision systems related to data development and algorithmic program norms, strengthening protection of personal privacy and information security. In May 2018, the EU promulgated the General Data Protection Regulation (GDPR), which clearly defined rights to data access, modification, erasure, and objection. Targeted laws and regulations can effectively regulate data mining and algorithmic programs, proactively constraining algorithm users' behaviors, representing an important approach to constructing algorithmic ethics.

While improving laws and regulations, relevant supervision standards and industry norms should be established to reasonably allocate weights for different content such as morality, education, art, and religion in practice, enabling AI design systems to coordinate and unify with human society's moral standards.

Second, construct internal and external algorithmic governance constraint mechanisms, enhancing value-sensitive design of algorithmic programs. Computer system algorithmic bias can be divided into pre-existing bias, technical bias, and emergent bias. "Pre-existing bias originates from social institutions, attitudes, and habits" [10]. For example, searching "CEO" on Baidu generally displays male figures. Technical bias mainly results from technological limitations on program development and design, while emergent bias arises from conflicts in social culture and value concepts, such as when Microsoft's AI chatbot Tay on Twitter learned discriminatory bias from users' racially discriminatory remarks.

For algorithmic program designers and developers, "we need to strengthen the construction of algorithmic designers' moral responsibility and improve the construction of moral imagination" [11]. During algorithmic development, possible consequences of various scenarios can be assessed through risk evaluation, and effective decision-making standards can be made through these assessments. Embedding moral concepts throughout the entire algorithm system development and application process, value conflicts can be balanced from quantitative and qualitative analysis perspectives.

Third, improve AI technological innovation capabilities and govern data cleaning issues according to scientific principles. "Developers should adhere to basic ethical norms, integrating principles of safety, beneficiality, and inclusive diversity throughout design, with clear responsibilities" [12]. AI systems should not be used for illegal or unethical purposes. Rights and obligations of subjects in AI design, R&D, operation, manufacturing, and service should be clarified, as should the responsibilities moral subjects should bear when ethical risks occur.

In the data collection phase, datasets should be screened, cleaned, and formatted according to industry standards to avoid algorithmic decision-making unfairness caused by data bias. In the data development phase, we should abandon data hegemony concepts of excessive data extraction and collect data moderately according to market mechanisms and industry standards. "Data development platforms should earnestly fulfill their main responsibilities, strengthen technical leadership, and promote self-discipline among internet giants" [13]. Internet enterprises should strengthen normative construction and technical protection for advertisers and user data reserves, management, and transmission, adhering to professional ethics, morality, and spirit in data use and advertising communication. This can not only protect enterprises' legitimate rights and interests but also safeguard user rights, and dutifully consolidate the maintenance of national network sovereignty, security, and development interests.

In the final application phase, "we should improve transparency in algorithmic design, reduce personal information leakage risks through technical support,

advocate for fair, open, and inclusive algorithmic mechanisms, and construct two-way interactive feedback mechanisms” [14]. This enables platforms to collect user information demands in real-time, forming a mutually beneficial and sustainable benign cycle.

References

- [1] Huang Heyan. Understanding and Reflection on Talent Cultivation for Artificial Intelligence Majors Under the Background of Emerging Engineering Education [J]. *China University Teaching*, 2019(2): 20-25.
- [2] Deng Guomin, Li Mei. Discussion on Ethical Issues and Principles of Educational Artificial Intelligence [J]. *e-Education Research*, 2020(6): 39-45.
- [3] Song Jixin, Wei Yudong, Wang Yongfeng. Review of Research on Big Data Ethics Issues and Governance [J]. *Theory Horizon*, 2017(1): 48-54.
- [4] Zhang Guihong. Value-Sensitive Design and Big Data Ethics [J]. *Ethics Research*, 2019(2): 114-119.
- [5] Li Lun. “Truman Effect” : The “Ideology” of Data Machines—Dataism and Rights-Based Data Ethics [J]. *Exploration and Free Views*, 2018(5).
- [6] Jia Honglin. Research on Big Data Algorithm Ethics [D]. Changsha: Hunan Normal University, 2020.
- [7] Li Lun. “Truman Effect” : The “Ideology” of Data Machines—Dataism and Rights-Based Data Ethics [J]. *Exploration and Free Views*, 2018(5).
- [8] Lv Yaohuai. Ethical Considerations of Information Security in the Big Data Era [J]. *Morality and Civilization*, 2019(4): 84-92.
- [9] Yuan Xue. The Ethical “Seven Deadly Sins” of Big Data Technology [J]. *Science and Technology Communication*, 2016(4).
- [10] Cheng Sumei, Zhang Fan. Philosophical Issues of Artificial Intelligence [M]. Shanghai: Shanghai People’s University Press, 2020: 185-186.
- [11] Duan Weiwen. Ethical Strategies for the Artificial Intelligence Era [J]. *Contemporary American Review*, 2019(1): 24-38+120.
- [12] Li Lun, Sun Baoxue, Li Bo. Ethical Constraints on Big Data Information Value Development: Mechanism Framework and China Focus [J]. *Journal of Hunan Normal University (Social Science Edition)*, 2018(1): 1-8.
- [13] Li Lun, Li Bo. Ethical Issues in Information Value Development in the Big Data Era [J]. *Ethics Research*, 2017(05): 100-104.
- [14] Yan Kunru. Moral Risks of Artificial Intelligence and Their Avoidance Paths [J]. *Journal of Shanghai Normal University (Philosophy and Social Sciences Edition)*, 2018(2): 40-47.

Author Profile: Yang Danxiu (1972-), male, from Shaoyang, Hunan, is an associate professor and doctoral supervisor at the School of Media and Communication, Yunnan University of Finance and Economics. His research focuses on intelligent communication.

(Responsible Editor: Zhang Xiaojing)

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.