DOI: 10.5281/zenodo.14569599
Vol. 2(3); 2024
Received 09 Oct 2024 ; Accepted 10 Dec 2024
Available online 30 December, 2024, Copy right @2024

Journal of Interdisciplinary Insights
ISSN (Online) :2995-6587
Published by Global Insight Publishing Ltd, USA

Legal Risks and Regulatory Frameworks for Artificial Intelligence

Liu Xinyu[1] , Yang Xinmiao[2] , Zhang Mingyang[3]

Abstract: The promise of innovation and practical applications of artificial intelligence has garnered significant interest and serves as a new motivator for societal advancement. The ensuing legal ramifications have attracted considerable examination. The legal risks associated with artificial intelligence development can be categorized into three areas: data security vulnerabilities during the training phase, algorithmic bias concerns in the generating phase, and the potential for misuse as illegal or criminal instruments in the application phase. This article employs a thorough research methodology, encompassing a literature review of notable legal cases and laws, alongside case-specific evaluations of particular artificial intelligence applications. Research findings indicate that a comprehensive regulatory framework, which meticulously assesses both the process and the outcome, is essential for mitigating legal risks at every level of artificial intelligence. This framework must encompass data security safeguards, mitigation of algorithmic bias, and explicit delineation of responsibilities. The investigation into the regulatory framework is crucial as it can mitigate legal challenges and foster social advancement by establishing a definitive legal guideline for the progression of artificial intelligence.
Keywords: Artificial Intelligence; Legal Risks; Regulatory Frameworks

Introduction

In recent years, artificial intelligence technology has achieved significant advancements and novel formats have emerged. This technology is progressively serving as a pivotal catalyst for scientific and technological innovation, economic advancement, and societal improvement, encompassing text generation, image synthesis, voice simulation, and video production. In February 2024, the American artificial intelligence firm OpenAI introduced their latest Wensheng video model, Sora, and unveiled a collection of video examples produced by this model. [1]. Sora's video production system demonstrates exceptional performance, capable of transforming text instructions or static images into one-minute high-definition videos, marking a significant milestone in the advancement of artificial intelligence.

The advancement of artificial intelligence technologies in China has also progressed significantly. As of August 2024, China boasted over 190 substantial artificial intelligence service models available to the public, with in excess of 600 million registered customers. [2]. The advancement of artificial intelligence technology represents a transition from conceptual validation to practical implementation, from laboratory prototypes to widespread industrial deployment. Prominent companies like Iflytek, Baidu, Alibaba, and Tencent have significantly reduced the accessibility barrier of this technology by offering user-friendly platforms and tools. In March 2023, little than 20 days following the launch of ChatGPT by Samsung Electronics, sensitive information was disclosed. [3]. In that year, the offense of utilizing Chat GPT emerged for the first time in China. A man utilizing Chat GPT's Deepfake technology to create fabricated news about a "train striking individuals" for the purpose of generating traffic may be subject to a five-year prison sentence. [4]. In February 2024, the Guangzhou Internet Court rendered a decision on the inaugural case of infringement involving a global artificial intelligence generation platform. [5]. The artificial intelligence painting feature on the defendant's platform has undergone extensive training and produced infringing images. The platform was deemed primarily responsible due to the flawed complaint reporting system, inadequate risk warnings, and absence of clear indicators. The aforementioned examples illustrate that the legal dangers associated with artificial intelligence technologies are becoming increasingly significant.

[1] (First Author) Crimial Investigation School, Southwest University of Political Science and Law, People's Republic of China.
[2] (Corresponding Author) Baize Institute for Strategy Studies, Southwest University of Political Science and Law, People's Republic of China.
[3] (Corresponding Author) Baize Institute for Strategy Studies, Southwest University of Political Science and Law, People's Republic of China.
Corresponding email: lirambo566@gmail.com (Liu Xinyu), 18545529458@163.com (Yang Xinmiao), 15684168158zmy@sina.com (Zhang Mingyang)

Literature Review

In 2021, UNESCO released the inaugural normative framework document regarding the ethics of artificial intelligence, titled "Recommendation on the Ethics of Artificial Intelligence", which was officially ratified by 193 member states. [6]. The International Organization for Standardization has commenced an evaluation of artificial intelligence security, while the World Digital Technology Academy published two international standards in April 2024: "Generative Artificial Intelligence Application Security Testing Standard" and "Large Language Model Security Testing Method". [7]. China has proactively enhanced the governance of artificial intelligence through both domestic and international legal frameworks, subsequently enacting legislation and administrative regulations including "Law on Scientific and Technological Progress" "Law on Network Security" "Law on Data Security", and "Law on Personal Information Protection". The Interim Measures for the Management of Generative Artificial Intelligence Services, the inaugural legislation on generative artificial intelligence jointly released by seven departments, including the National Internet Information Office, became effective in August 2023. [8]. Furthermore, China engages in global artificial intelligence governance, with the Ministry of Foreign Affairs releasing "Global Artificial Intelligence Governance Initiative" in October 2023, promoting the principle that artificial intelligence technology development should be "people-oriented". [9].

Academic circles are significantly concerned about the legal hazards associated with artificial intelligence and have generated numerous study findings of practical use. Nonetheless, the inherent traits of its nascent technology result in persistent delays and constraints in theoretical study within this domain. The risk mechanism of this technology has not been thoroughly investigated, limiting the development of more targeted regulatory initiatives. The current proposed regulatory policies are restricted in scope, mostly concentrating on certain domains such as intellectual property rights. This paper will thoroughly examine the mechanisms of legal risks associated with artificial intelligence, encompassing its technical support, application domains, and potential modes of misuse. It will also assess the efficacy and limitations of the current legal framework in addressing these risks, aiming to explore more comprehensive, multi-tiered, and multidimensional regulatory strategies.

Methodology

This research aims to thoroughly examine the legal hazards associated with artificial intelligence and is dedicated to developing an appropriate regulatory framework. This paper utilizes a comprehensive and rigorous research methodology to attain the research objective.

Regarding the literature analysis process, legal case studies are conducted. Systematically and extensively gather a substantial array of pertinent legal cases pertaining to artificial intelligence, both domestically and internationally, encompassing issues relating to data security, algorithmic discrimination, and the illicit utilization of artificial intelligence. Thoroughly examine their particular circumstances, including the event's context, the legal dispute's focal points, the ultimate verdict, and the ensuing ramifications. For instance, perform a comprehensive investigation of the incident involving Samsung Electronics' utilization of ChatGPT, which led to the breach of confidential information, and the case addressed by the Guangzhou Internet Court over the infringement of an artificial intelligence generation platform. Identify the essential elements pertaining to the legal hazards associated with artificial intelligence from these instances to establish a solid foundation for later risk categorization and regulatory framework development. Conversely, undertake an investigation of statutes and regulations. Thoroughly and methodically categorize the normative documents pertaining to artificial intelligence issued by international organizations (e.g., UNESCO, ISO), including the "Recommendation on the Ethics of Artificial Intelligence" published by UNESCO. Additionally, encompass the various laws and regulations concerning artificial intelligence enacted by different nations (notably China), such as "Law of the People's Republic of China on Scientific and Technological Progress" "Cybersecurity Law of the People's Republic of China" "Data Security Law of the People's Republic of China" "Personal Information Protection Law of the People's Republic of China", and " Interim Measures for the Management of Generative Artificial Intelligence Services" among others. Thoroughly examine the specific clauses, legislative intents, and applicability of these laws and regulations to elucidate the legal requirements concerning various facets of artificial intelligence, thereby establishing a robust theoretical basis for assessing the efficacy of the existing legal framework in addressing the risks associated with artificial intelligence.

Concerning the case evaluation methodology, concentrate on the assessment of certain artificial intelligence applications. Select representative artificial intelligence applications for comprehensive review, encompassing diverse categories such as text generation, image synthesis, voice simulation, and video production. The assessment encompasses applications like OpenAI's GPT series and Midjourney. For each chosen application, meticulously analyze every link in its operational process. In the data collection link, meticulously examine the source and methodology of the application's data acquisition, while also assessing compliance with pertinent laws, regulations, and ethical standards; during the model training phase, thoroughly investigate its training algorithm, data utilization, and the security measures implemented for data protection; regarding the output of the generated content, critically

appraise its quality, accuracy, and the potential for discriminatory or unlawful content; in the application context, attentively observe the user's application usage, the possible ramifications of the application, and any risks of misuse.

Results

According to "Black's Law Dictionary", regulation is defined as the act or process of controlling through rules or restrictions, which signifies that regulation encompasses the creation and enforcement of rules, standards, or limitations, while the objective of regulation is to govern a particular action or procedure, ensuring compliance with legal, ethical, and social standards. To successfully address the legal concerns associated with artificial intelligence, it is essential to create a comprehensive regulatory framework that combines process control and outcome regulation. Guided by this paradigm, we may systematically mitigate risks across three stages, ensuring the robust development of technology while safeguarding user rights and public interests.

Standardizing the Concept of Development: Compound Supervision Path

Risk regulation frequently necessitates decision-making in uncertain circumstances. Uncertainties exist regarding the occurrence, timing, and nature of artificial intelligence risks, necessitating that artificial intelligence law prioritize risk prevention as a fundamental principle and advocate for comprehensive regulation of these risks to achieve proactive governance. [10]. The fundamental concept of standardizing the development trajectory must be dynamic and future-oriented, thoroughly addressing the increasing risks, including the entirety of legal risk management, and directing the overall strategy and implementation of the activity. The optimal normative growth trajectory must incorporate both process regulation and result regulation tactics, employing a multifaceted regulatory approach to comprehensively manage risks. [11]. Procedural regulation is an adaptive and engaging approach, whereas consequential regulation is a fixed and retributive strategy implemented subsequently.

Procedural regulation emphasizes real-time oversight and direction of all components throughout the entire process, with the monitoring reach extending beyond mere outcomes. Procedural regulation mandates that all stakeholders, including creators and users of artificial intelligence technology, adhere to explicit operational protocols and standards, and consistently revise them to ensure their relevance. Encompassing data collection, processing, storage, application, algorithm creation, testing, and deployment. Implementing a transparent and traceable process management system enables the creation of a stable technological ecosystem and facilitates early intervention to mitigate potential legal issues at their origin. Results-based regulation emphasizes the assessment and evaluation of the outcomes of certain actions or processes to guarantee that the results adhere to defined legal, ethical, and social norms. When the outcome fails to meet standards, the regulatory framework mandates the accountability of the responsible entity. Consequential regulation emphasizes the quality and impact of produced synthetic content, while overseeing the specific application effects. The method of post-intervention may result in a delay in addressing adverse effects. They both fulfill complimentary functions in regulation, and their combined application can thoroughly oversee the decision-making process, algorithmic bias, content output, and potential harm.

Guarantee Data Security and Uphold the Right to Information Control

The conventional personal control theory of personal information relies on individualism but neglects the social and public dimensions of personal information, rendering it inadequate for the evolving landscape and novel methods of personal information usage in the era of big data. Information protection must transition from individual oversight to collective governance. [12]. The capacity for information control is essential for safeguarding data security and personal privacy, while the rights to know and consent constitute the foundation for ensuring individual information self-determination. [13]. The functioning of artificial intelligence encompasses nearly the entire spectrum of personal information processing, as defined in Article 4 of Personal Information Protection Law of the People's Republic of China, which includes the collection, storage, utilization, processing, transmission, provision, disclosure, and deletion of personal information. [14]. More comprehensive legislation and guidelines are required at the legal level to ensure the effective protection of information rights and interests. This article proposes two regulation recommendations regarding the conduct of artificial intelligence in acquiring and utilizing data.

The primary objective is to enhance the transparency and equity of the process for gathering human data driven by artificial intelligence. The right to know grants people access to the operational mechanisms, decision-making logic, and automated processes of artificial intelligence systems that may impact their rights and interests. [15]. The right to consent necessitates that you make a decision about the acquisition and utilization of your personal information after comprehensively considering the aforementioned information. The transparency and equity of the instructional process for gathering human data by artificial intelligence underpin the rights to knowledge and consent. Article 16 of Regulations on the Management of Algorithmic Recommendations for Internet Information Services mandates that providers of algorithm recommendation services must prominently inform users about these services, including the fundamental principles, objectives, intentions, and primary operational mechanisms involved. Prior to gathering user data, the developer must secure the user's explicit consent and provide information regarding the data collection methods, usage, processing, source, purpose, retention duration, and any potential data sharing. Developers must not employ deceptive or misleading tactics to secure users' consent; consumers should be informed of this information and possess the genuine option to consent or decline, as well as the ability to request data deletion or rectification.

The second objective is to set guidelines for anonymization and de-identification to assure data reduction and prioritize the use of synthetic data. China's excessive quest for anonymity, the absence of legal efficacy of rules, and the fragmented legislative resources impede the proper utilization and safeguarding of personal information. [16]. To guarantee the data security of artificial intelligence, it is essential to develop specific standards and procedures for data anonymity and de-identification tailored to various data types and application scenarios, while instituting the principle of data minimization and promoting the preferential use of synthetic data. Eliminate, modify, or handle personally identifiable information to avert its association with particular individuals. The notion of data reduction mandates the collection of only the data essential for accomplishing a certain job. Reducing data volume can streamline data management, lower storage and processing costs, enhance control over data flow and utilization, and mitigate the risk of data breaches. Simultaneously, artificial intelligence training should prioritize synthetic data produced by the algorithm that emulates the attributes of real data while excluding actual individual information. Synthetic data offers secure and dependable training and testing datasets for the model while significantly mitigating the danger of data misuse. Utilizing face recognition technology, reliance on actual personal data can be circumvented by employing a synthetic face picture training model.

Mitigate Algorithmic Bias and Enhance Trust in Algorithmic Functionality

The risk mechanism during the generation and synthesis stage indicates that the algorithm's discrimination risk primarily arises from the deviation or imbalance in training data, along with issues related to model structure and parameter configuration. Consequently, we can mitigate algorithmic bias by optimizing data sets, enhancing models and parameters, refining specifications, and bolstering oversight.

The initial step is to thoroughly gather data, equilibrate data categories, and enhance data sets. Enhancing the dataset for artificial intelligence training is essential for algorithm functionality and superior output quality. The dataset must encompass a wide array of circumstances and scenarios to guarantee that the model acquires sufficient knowledge from it. Diversity enhances the model's ability to generalize, mitigates overfitting, and elevates its performance in real-world applications. Furthermore, due to the disparity in sample sizes across several categories, techniques such as resampling, oversampling, or undersampling may be employed to equilibrate the sample distribution throughout the categories. In the context of medical image recognition, the quantity of images depicting normal cells significantly exceeds that of diseased cells. The model exhibits significant bias towards several categories throughout training, resulting in a substantial decline in its recognition capability for a limited number of categories.

The second objective is to enhance model transparency, diminish algorithm design subjectivity, and refine structure and parameter configuration. Visual tools can be employed by technical teams to present intricate algorithmic processes and substantial data in a simple and comprehensible manner, hence enhancing the transparency of artificial intelligence. For instance, TensorBoard and Netron can display the input and output of each layer of a neural network, as well as the alterations in the model's weights during training. [17]. Simultaneously, it is essential to enhance the diversity of the algorithm design team, mitigate or at least diminish biases stemming from a singular perspective, and the involvement of an external team can also provide greater objectivity to the review process. At the institutional level, we can reference the regulatory exclusive right system to safeguard drug data, granting the algorithm developer market exclusivity for a specified duration in exchange for the disclosure of the algorithm (excluding the source code) to address the challenge of supervising the algorithm. [18]. A governance paradigm for public participation algorithm discrimination, grounded in deliberative democracy, is established. Public involvement in the control of algorithmic discrimination aligns with Finberg's democratic concept of "design criticism". Public involvement in the regulation of algorithmic discrimination directly addresses the upstream aspects of algorithm development and auditing, thereby embodying democratic attributes. [19].

The third objective is to enhance technical requirements and augment external oversight. The current normative papers offer broad recommendations for rectifying algorithmic bias; nonetheless, there remains an absence of particular implementation and detailed operational requirements. This industry must enhance its rules and set definitive technical standards. Simultaneously, due to the interdisciplinary nature of generation technology, it is essential to delineate the primary regulatory bodies for artificial intelligence and eliminate the regulatory gaps resulting from several departments. Secondly, it is essential to enhance the coordination and collaboration among the National Network Information Office, the Ministry of Industry and Information Technology, and other departments to establish an effective regulatory authority. Enhance the capacity to examine and penalize algorithmic discrimination, guarantee the swift resolution of infractions, and implement appropriate fines and accountability frameworks. Furthermore, a robust user feedback system should be implemented to bolster the safeguarding of users' rights and interests. A body like to the Consumers Association may be formed to promote the active reporting of bias or discrimination by users.

Subdivide the application context and delineate the subject of liability

The prevailing perspective asserts that artificial intelligence lacks independent consciousness and volition, remaining an auxiliary tool for humanity. Within the domain of cognitive research, human cognition is categorized into five tiers: neurological, psychological, linguistic, cognitive, and cultural. Artificial intelligence operates primarily by emulating human cognitive processes at both the brain and psychological levels. It is fundamentally a reductive replica that inadequately represents the intricacies and fluctuations of human cognition. There exists a fundamental distinction at

the linguistic level between the artificial language employed by artificial intelligence and human natural language, with the former frequently lacking the profound meaning and contextual nuance inherent in the latter. At the cognitive and cultural levels, contemporary artificial intelligence technology has yet to demonstrate genuine cognitive ability or cultural originality. [20].

The existing Criminal Law of China does not acknowledge artificial intelligence as an autonomous subject of criminal liability. Criminal activities including artificial intelligence fundamentally remain human criminal behaviors. [21]. When artificial intelligence operates according to its programmed parameters, as a tool or an infringing entity, the accountability typically resides with the developer or user. The accountability system for artificial intelligence can be hierarchically structured based on the risk associated with the artificial intelligence system, following the principle of imputation. Different levels of risk may necessitate the adoption of no-fault liability, fault assumption liability, or general fault liability principles. [22].

Nevertheless, the growing prominence of intelligence and humanoid traits in artificial intelligence has begun to contest the conventional algorithm as a mere tool, highlighting its potential as a subject of study. In the future, artificial intelligence may acquire Theory of Mind capabilities, enabling it to comprehend and reason human intents, beliefs, and emotions, so demonstrating a level of understanding comparable to that of adults. The current framework that perceives artificial intelligence only as a tool and object, while imposing all legal liabilities on developers through penetration algorithms, may soon encounter significant obstacles. [23]. It is anticipated that as algorithmic intelligence and human-like traits advance, artificial intelligence will go from computational intelligence to perceptual intelligence, ultimately achieving cognitive intelligence characterized by autonomous cognitive capabilities. In the realm of computational intelligence, the algorithm predominantly exemplifies data processing capabilities and possesses distinct tool characteristics. During the perceptual intelligence phase, algorithms are embedded within particular contexts to assist individuals in decision-making, demonstrating their dual characteristics as tools and products.

There is an urgent need for additional theoretical discourse and demonstration to examine the criminal liability of advanced artificial intelligence robots that exceed programmed autonomous conduct. It necessitates a comprehensive examination of behavioral autonomy, decision-making processes, and the potential ramifications of artificial intelligence, while also requiring a thorough consideration of the adaptation and response to the swift advancement of technology within the legal framework. When artificial intelligence operates independently of programmed actions, and both the creator and user lack criminal intent and have adhered to their duty of care, it may be deemed appropriate to hold the artificial intelligence itself independently accountable. When the developer or user exhibits criminal intent or breaches the duty of care, and the artificial intelligence independently engages in accomplice activity outside program control, the participants may be seen to share blame.

Discussion

Ulrich Beck's theory of risk society posits that the rapid expansion of productive forces throughout modernization results in an unparalleled proliferation of risks and potential self-hazards. [24]. Generative artificial intelligence is a technology that employs Deep Learning and Machine Learning algorithms to identify patterns and features within large datasets, comprehend the generative principles of content or objects, and subsequently autonomously produce novel texts, images, music, and videos that are not directly created by humans. In contrast to traditional artificial intelligence, which primarily focuses on passive data analysis and model recognition, contemporary "Generative AI" possesses the capability to actively generate content and make decisions through self-synthesis. This technology encompasses various values, including informational, emotional, cognitive, and labor-related aspects, and is anticipated to evolve into Artificial General Intelligence (AGI), potentially leading human society into a landscape characterized by high risk and high reward.

The legal risks associated with artificial intelligence exhibit distinct technical attributes in contrast to traditional legal risks. The technology evolves rapidly, while the legal response lags behind. Despite the inherent delay of legal frameworks, the rapid advancement of artificial intelligence technology outpaces legislative adaptation. Consider GPT-4 and GPT-4O, which were introduced by OpenAI merely one year apart. [25]. [26]. The latter has significantly enhanced the processing of non-English text, increased response speed, reduced application programming interface (API) costs, and enabled the acceptance of any combination of text, audio, and images as input to produce comparable output. The technical distinctions between the two products are substantial. Secondly, the utilization of technology is concealed, presenting challenges for legal recourse. The output of artificial intelligence technology is challenging to differentiate from authentic information, complicating legal oversight. In 2023, a Chinese person employed ChatGPT to concoct a deceptive news item regarding the annulment of the motor vehicle restriction policy in Hangzhou, resulting in widespread public misperception of its veracity. [27]. In that year, political interference utilizing artificial intelligence significantly influenced Slovak parliamentary elections, and an audio recording alleging that political elites were conversing about election manipulation circulated extensively on social media. [28]. Third, technology encompasses numerous disciplines, is cross-regional, and presents issues of ambiguous legal accountability. The occurrence of damage implicates multiple parties, including developers, consumers, and service providers, rendering the attribution of culpability complex. Furthermore, the aforementioned subjects are not constrained by geographical

borders, and cross-regional interactions are evident, making the assignment of responsibilities challenging. The robust technical attributes of artificial intelligence legal risk need an examination of its risk mechanism from a technical perspective.

The essential approach to examining the legal risk framework of artificial intelligence from a technological perspective is to scrutinize the precise procedures involved in the intelligent decision-making process. Artificial intelligence depends on the aggregation of extensive data and the development of algorithms, thoroughly examining and analyzing existing information, comprehending the statistical properties of particular content types, and subsequently interpreting and replicating intricate data distribution patterns, thereby achieving precise mapping from data input to output results. The entire procedure, encompassing the pretreatment of original data, the development of new data, and the ultimate application, may be categorized into three essential temporal stages: the training stage, the generation and synthesis stage, and the application stage. Each phase is associated with particular dangers, specifically, data security risk, algorithmic discrimination risk, and the potential of evolving into an illicit and criminal instrument.

Training Phase: Data Security Risks

"The Cost of a Data Breach Report 2024" indicates that the average expense of a data breach in 2024 reached 4.88 million US dollars, marking a record high and a 10% increase from 2023. With the advancement of cloud computing, the Internet of Things, and other technologies, the generation, transfer, and storage of data are increasingly prevalent, rendering current data a significant asset. Numerous countries and regions have implemented specific data protection laws, such "California Consumer Privacy Act" (CCPA) and "Personal Information Protection and Electronic Documents Act" (PIPEDA) of Canada, which provide comprehensive standards around data collection, storage, processing, and utilization. China has instituted a comprehensive array of legislation and regulations pertaining to data security protection.

Numerous issues exist in the current implementation process. The provisions are predominantly principled and generic in nature. Despite "Personal Information Protection Law" of China and similar regulations proposing the "necessary principle", varying views on its precise understanding and implementation persist, potentially resulting in network operators excessively collecting user information while delivering services. Furthermore, it is mandated that users be explicitly informed regarding the goal, methodology, and extent of personal information gathering and utilization, and obtain users' consent. Users frequently encounter the predicament of "agree or forgo use", undermining the efficacy of the user consent principle.

During the training phase, the system will acquire various original data, user information, feedback, and directives, among other inputs. The efficacy and security of the model are directly correlated with data processing and learning. The training phase encompasses data pre-processing, model training, optimization, and fine-tuning. The initial phase is Data Preprocessing. The system must filter and cleanse the original data, remove noise and redundant information, and retain only the valuable components for model training. Subsequently, through Feature Extraction and Word Vector Encoding, the data undergo normalization and Min-Max Scaling, resulting in a unified data format and adjusted data range, therefore mitigating scale discrepancies among various feature values and enhancing the algorithm's stability and efficiency. Following data preparation, which establishes the groundwork for Neural Networks' subsequent learning, the second step—feature learning—is conducted to develop the model. A neural network emulates the properties of human brain neurons to assimilate input data and develops a data representation model that can encapsulate the intricate structure and patterns of data, establishing a basis for the generation phase. Ultimately, the model's optimization. Artificial intelligence incrementally minimizes the discrepancy between the model's predictions and the actual outcomes by modifying the network weights. By computing the gradient of the Loss Function with respect to the model parameters and adjusting the parameters in the opposite direction as the gradient, the value of the loss function is minimized, leading to the identification of the best solution. The system will concurrently gather and address instructions and feedback from human users, then transforming them into data for refining the pre-trained model, ultimately producing high-quality output that satisfies user requirements.

Data preprocessing and model optimization entail data absorption, which may be a possible gateway to data security vulnerabilities. The data accessed by the system may include personal privacy information or content safeguarded by intellectual property rights. Utilizing data from non-public channels may contravene laws and regulations if employed without the agreement of the data subject or in compliance with applicable legal standards. When handling sensitive data, including identity information, personal preferences, behavioral patterns, and biological traits, direct utilization for model training without adequate anonymization or de-identification may result in the reemergence of such data in subsequent generations and syntheses, thereby posing risks of privacy breaches and violations. Although the data originates from open sources, the limitations on its commercial usage require more deliberation. [29]. During model optimization, human-computer interaction occurs, enabling the system to enhance task performance by direct user instructions. During human-computer contact, there exists a danger that users may unintentionally reveal or be prompted to gather information. Artificial intelligence can tag and rearrange fragmented information, enabling the construction of comprehensive user profiles and the extraction of sensitive data.

Generation and Synthesis Phase: Algorithm Discrimination Risks

During the production and synthesis phase, the neural network produces novel data that adheres to a particular

distribution or principle based on the information and experience acquired during the training phase. The essential technology at this juncture is the generation of Generative Adversarial Networks (GANs), which operate on the notion of competition between the Generative Model and the Discriminative Model. The former is tasked with producing authentic data samples, whereas the latter aims to differentiate between genuine samples and those manufactured, ultimately developing a countermeasure network to progressively enhance the quality of the generated data.

The possibility of algorithmic prejudice is significant during data creation and synthesis. The system categorizes or forecasts persons based on specific attributes, perhaps resulting in inequitable outcomes. If the algorithm employed for public safety, law enforcement, or judicial assessment relies solely on historical data or correlation training, individuals may be erroneously classified as high-risk groups despite having no actual wrongdoing. [30]. The fundamental premise of algorithmic discrimination is that the generator preferentially produces specific data kinds while neglecting others, or the discriminator exhibits bias towards certain data types, leading to erroneous assessments. Algorithmic discrimination renders information singular and biased, exacerbating the Echo Chamber Effect. Research indicates that artificial intelligence-guided self-driving cars exhibit deficiencies in identifying dark-skinned individuals, potentially endangering their safety. [31]. The primary factors contributing to generator bias or discriminator bias often pertain to training data, model architecture, and parameter configuration.

The quality of the training data directly influences the final output. If the training data is skewed or distorted, the generated data will eventually replicate this bias, resulting in discriminatory output. The predictive policing model, developed using past crime data, exhibits disproportionate surveillance of a specific demographic, potentially unjustly categorizing this group as high-risk. An imbalance in the training data may cause the generator to favor a certain kind, so restricting its ability to generate data for other categories or groups. Concurrently, the discriminator will exhibit bias against this category of data. For instance, if the training data predominantly emphasizes the faces of young individuals or particular ethnicities, the model may preferentially produce such faces, neglecting other age demographics or ethnic groupings.

The irrational model structure and parameter configuration result in bias or erroneous conclusions. Inadequate model architecture and parameter configuration will result in the created countermeasure network's inability to effectively capture intrinsic data properties and patterns. Nevertheless, the model comprises a substantial quantity of free parameters. Consider the AlexNet model, a preeminent framework in the domain of image recognition. The model, which comprises more than 62 million free parameters, has its core data processing flow and algorithmic mechanisms, such as feature extraction from input data, reasoning, and output generation, remaining opaque to both end users and even developers. [32]. Currently, it is challenging to identify which neurons influence the specific elements of the output results and to measure the interaction between the two generative adversarial networks. [33]. A multitude of neurons and hierarchical structures inside neural networks are interconnected and interact through intricate weights and parameters, which are convoluted and obscure. The selection of parameters typically relies on experience, trial and error, and extensive experimentation, lacking a definitive theoretical foundation and rationale. Furthermore, the subjectivity of algorithm designers is exacerbated by the absence of diverse teams and insufficient oversight mechanisms. The designer inadvertently incorporates personal background and experiential biases into the algorithm, resulting in model outputs that favor the designer's subjective perspective rather than objectively representing the entirety of the data.

Application Stage: Risk of Being Exploited for Illicit Activities

The Department of Homeland Security's (DHS) "2024 Homeland Threat Assessment (HTA)" identifies "Foreign Misinformation" as one of the four principal risks to homeland security, alongside "Foreign and Domestic Terrorism" "Border and Immigration Security", and "Economic Security". [34]. The paper indicates that the nation-state may employ a strategy of disseminating misleading and erroneous information, utilizing network and artificial intelligence tools to conduct negative influence operations. Professional artificial intelligence models, including Midjourney, DALL-E, and Stable Diffusion, can produce visuals that may mislead viewers based on the input provided by users. It utilizes deep forging technology to embed bogus information into original audio and video sources using algorithms, rendering it highly convincing. In April 2023, Eliot Higgins, the founder of the open-source investigative media platform "Belling Cat", utilized the advanced artificial intelligence painting tool Midjourney to create a fabricated image of former US President Trump being subdued by heavily armed riot police in New York, subsequently sharing the manipulated results on Twitter. [35]. Despite the prior introduction of the DEEP FAKES Accountability Act and associated rules by the U.S. Congress, its impact was little.

Emerging domains such as Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), and computer vision have catalyzed significant advancements, facilitating the commercialization and everyday use of artificial intelligence. The authority to improve artificial intelligence has been broadened beyond the scientific team, allowing ordinary users to facilitate the system's self-optimization and iteration through practical application. Nevertheless, in the absence of requisite professional expertise and ethical limitations, technology may be employed for inappropriate purposes, either inadvertently or deliberately, and may even serve as an instrument for criminal activities. Utilizing deep forgeries technologies to effortlessly produce realistic audio and video, fabricate and disseminate fake material, harm the reputation and personal rights of individuals, and perhaps propagate terrorist

content. Employ machine learning to devise more nuanced fraud techniques, examine user activity patterns, and execute targeted fraud. Unauthorized reproduction and distribution of copyrighted materials, replication and counterfeiting of patented items, and infringement upon the rightful rights and interests of original artists via automated plagiarism and duplication technologies. Furthermore, the illicit use of artificial intelligence is continually enhancing the organization, specialization, and concealment of criminal activities. The illicit use of artificial intelligence has progressively transformed from a loosely organized group into a structured corporate crime with a defined division of work. [36].

Conclusion

Artificial intelligence depends on the aggregation of extensive data and the development of algorithms to thoroughly extract and analyze existing information, achieve comprehension, facilitate learning, adapt to the environment, and execute intelligent activities. Its evolution has progressed swiftly from single language creation to multi-modal and embodied forms. Artificial intelligence technology is advancing swiftly as a pivotal representation of emerging productivity. Nonetheless, the risks associated with data security during the training phase, algorithmic discrimination during the production and synthesis phase, and potential misuse as a tool for illicit activities during the application phase must not be overlooked. It is imperative to maintain the principle of compound regulation, which thoroughly evaluates both the process and outcome, guarantees data security, and honors the right to information management. Simultaneously, it mitigates algorithmic bias and bolsters the algorithm's reliability. Subdivide the application situations and explicitly delineate the tasks. Despite this study examining the hazards associated with artificial intelligence operations and proposing relevant countermeasures and recommendations, comprehensive quantitative analysis and empirical research on these risks and ideas remain insufficiently conducted. Future study will focus on addressing deficiencies, enhancing the comprehension of legal hazards, and refining coping techniques through comprehensive data gathering and interdisciplinary empirical studies.

In his congratulatory letter to "2024 World Intelligence Expo", President Xi Jinping emphasized: "China is willing to work with countries around the world to grasp the new trends of the digital era, deepen international exchanges and cooperation in the digital field, promote innovation and development in the intelligent industry, accelerate the building of a cyber community with a shared future, and work together to create an even brighter future." [37]. The regulation of artificial intelligence has emerged as a pivotal concern in global governance, with China's active involvement and contributions being essential. Through the development and continuous enhancement of its regulatory framework, China can proficiently safeguard the interests of the nation and its citizens, while also offering insights and solutions for global artificial intelligence governance, fostering the establishment of a just and equitable international governance system for artificial intelligence, and advancing the welfare of humanity as a whole.

Author Contributions: Conceptualization, methodology design, software, formal analysis, investigation, resources, writing—original draft preparation, writing—review and editing, validation, Liu Xinyu; Corresponding Author, Yang Xinmiao, Zhang Mingyang. All authors have read and agreed to the published version of the manuscript.

Funding: This research is the phased achievement of the major project on the construction of the Chengdu-Chongqing economic circle of the Chongqing Social Science Planning (Project No.: 2023ZDSC02), the key project of the legal speciality of the Chongqing Social Science Planning/Chongqing Law Society (Project No.: 2023FX14), the major project of the Sichuan Provincial Philosophy and Social Science Fund (Project No.: SCIJ24ZD74), the key project of Sichuan University of Science & Engineering (Project No.: ZYB(K)-2024-03), and the general project of Sichuan Police College (Project No.: ZHKFYB2305).

Data Availability Statement: All of the data utilised in this investigation was obtained from publicly accessible sources. The following are the links to specific datasets: The CNKI platform is the source for Chinese papers, while the Google Scholar platform is the source for foreign papers and books. For electronic materials, please consult the links in the References section for specific acquisition platforms. No new data were generated during the research procedure. In the interim, there were no instances in which data was unavailable due to privacy or ethical constraints, as all of the data utilised was public. We are dedicated to ensuring that data remains accessible and transparent in order to facilitate the further exploration and reproducibility of academic research.

References

[1]. Shanghai Securities News. (2024, February 19). Sora's Emergence: Artificial Intelligence Will Lead a New Round of Industry Transformation. Retrieved August 17, 2024, from https://paper.cnstock.com/html/2024 - 02/19/content_1877896.htm

[2]. China Digital Economy Network. (2024, August 14). The Cyberspace Administration of China: There are more than 190 generative AI large models that have been filed and put into service in China that can provide

services. Retrieved September 11, 2024, from https://www.digitalchina.gov.cn/2024/xwzx/szkx/202408/t20240814_4875554.htm

[3]. The Paper. (2023, April 14). Is Samsung Considering Banning ChatGPT? Confidential Content Entered by Employees Will Be Transmitted to External Servers. Retrieved January 3, 2024, from http://www.thepaper.cn/newsDetail_forward_22568264

[4]. The Paper. (2023, May 12). Using AI to Fabricate Fake News of "Train Hits People"! The First ChatGPT Crime in China, a Man in Dongguan May Face Five Years in Prison. Retrieved June 8, 2024, from https://www.thepaper.cn/newsDetail_forward_23040223

[5]. First Instance Civil Judgment (2024) Yue 0192 Minchu No. 113 of Guangzhou Internet Court of Guangdong Province.

[6]. Institute of Science and Technology Strategy Consulting, Chinese Academy of Sciences Network.(2022, June 6). UNESCO Releases the "Recommendations on the Ethics of Artificial Intelligence". Retrieved August 1, 2024, from: https://casisd.cas.cn/zkcg/ydkb/kjzcyzxkb/2022/zczxkb202202/202206/t20220606_6458039.html

[7]. China National Radio Network Science and Technology Channel. (2024, April 17). The United Nations Science and Technology Conference released two security standards for large models, and Ant Group took the lead in one of them. Retrieved August 19, 2024, from: https://tech.cnr.cn/techph/20240417/t20240417_526669258.shtml

[8]. Xinhuanet. (2023, August 16). The Interim Measures for the Management of Generative AI Services came into effect. Retrieved August 17, 2024, from: http://www.news.cn/2023 - 08/16/c_1212256586.htm

[9]. Ministry of Foreign Affairs. (2023, October 20). Global Artificial Intelligence Governance Initiative. Retrieved August 17, 2024, from https://www.fmprc.gov.cn/web/ziliao_674904/1179_674909/202310/t20231020_11164831.shtml

[10]. Hu, X. W., & Liu, L. (2024). The Full Process Regulatory Logic and Institutional Response of Artificial Intelligence Risks. *Study and Practice*, 5, 22 - 30.

[11]. Wang, Q. Y., & Wan, G. H. (2023). The Dual Legal Regulatory Logic of Generative Artificial Intelligence. *Studies on Socialism with Chinese Characteristics*, 4, 72 - 84.

[12]. Gao, F. P. (2018). Personal Information Protection: From privacy protection to Social Control. *Chinese Journal of Law*, 40(3), 84 - 101.

[13]. Shang, X. X. (2022). The Functional Positioning and Institutional Application of Citizens' Civil Personal Information Rights. *Zhejiang Social Sciences*, 7, 32 - 40, 157.

[14]. Meng, F. Q., & Wang, Z. Y. (2024). The Legal Risks and Governance of ChatGPT from the Perspective of Personal Information Protection. *Media*, 3, 51 - 54.

[15]. Zhang, Y. Z. (2024). On the Realization of the Principle of Transparency of Artificial Intelligence by Law. *Journal of Political Science and Law*, 2, 124 - 137.

[16]. Jiang, B. (2024). Improving Data Security Governance and Promoting the Development of Artificial Intelligence Industry. *Journal of Information Security Research*, 10(8), 776 - 779.

[17]. Kahng, M., Andrews, P. Y., Kalro, A., & Chau, D. H. (2018). ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 88 - 97.

[18]. Liang, Z. W. (2020). On the Exclusive Right of Algorithm: A New Choice for Correcting Algorithmic Bias. *Political Science and Law*, 8, 94 - 106.

[19]. Chen, Q. Q., & Zhang, B. Y. (2024, June 19). Public Engagement in Algorithmic Discrimination Governance: A Study of the Democratizing Technology. *Studies in Science of Science*, 1 - 11.

[20]. Zhen, H. (2024). The Negation of the "Subjectivity" of Artificial Intelligence in Criminal Law: Origins, Deconstruction and Reflection: Based on the Five-Level Theory of Cognitive Science. *Journal of Chongqing University (Social Science Edition)*, 30(3), 242 - 252.

[21]. Fang, H. Y. (2022). The Doctrinal Expansion of the Attribution and Identification of Criminal Liability for Artificial Intelligence Crimes. *Shandong Social Sciences*, 4, 142 - 148.

[22]. Zheng, Z. F. (2024). Subject Identification and Imputation Design of Artificial Intelligence Application Liability. *Law Review*, 42(4), 123 - 137.

[23]. Zhang, X. (2023). Algorithmic governance Challenges and Governance Supervision in Generative Artificial Intelligence. *Modern Law Science*, 45(3), 108-123.

[24]. Beck, U. (2018). *Risk Society: Towards a New Modernity* (Zhang Wenjie & He Bowen, Trans.). Nanjing: Yilin Press.

[25]. The Paper. (March 15, 2023). The release of GPT-4, a "bombshell": Its professional and academic levels are close to those of humans, and it only takes 1 second to build a website. Retrieved June 8, 2024, from https://www.thepaper.cn/newsDetail_forward_22302931

[26]. The Paper. (May 15, 2024). OpenAI has launched the latest large model "GPT-4o", which can understand your joys and sorrows. Retrieved June 8, 2024, from https://www.thepaper.cn/newsDetail_forward_27372275

[27]. Guangming Online. (2023, February 17). Stop spreading it! This piece of fake news was written by ChatGPT and the police have already launched an investigation. Retrieved August 14, 2023, from https://m.gmw.cn/2023-02/17/content_1303287072.htm
[28]. Wittschafter, V. (2024, Issue 97). The Impact of Generative Artificial Intelligence in the Year of Global Elections. Overseas Think Tank Views Digest. Retrieved October 6, 2024, from http://en.iiss.pku.edu.cn/__local/3/7F/3D/9EA4A3B30950C7A22DAC5E56962_861E8C97_60B97.pdf
[29]. Lyu, Y. Y. (2023). Forward-looking on the Crime Governance of ChatGPT Technology. *Chinese Rule of Law*, 4, 56 - 60.
[30]. Li, S. (2023). Advantages and Hidden Dangers of Digital Governance. *Study & Exploration*, 10, 58 - 67.
[31]. PwC. (2023, December 4). Understanding algorithmic bias and how to build trust in AI. Retrieved September 4, 2024, from https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-bias-and-trust-in-ai.html
[32]. Ma, A., & Song, Y. Z. (2022). The Phenomenon of "Algorithmic Discrimination" in Artificial Intelligence Crime Risk Assessment and Its Regulatory Path. *Jianghuai Tribune*, 2, 119 - 127, 193.
[33]. Jovanović, M., & Campbell, M. (2022). Generative artificial intelligence: Trends and prospects. *Computer*, 55(10), 107 - 112.
[34]. US Government News. (2023, September 14). DHS releases 2024 Homeland Threat Assessment (HTA). Retrieved August 1, 2024, from https://usgovernmentnews.com/dhs-releases-2024-homeland-threat-assessment-hta/
[35]. Science and Technology Daily. (2023, April 3). The Proliferation of Fake Images Generated by AI Urgently Needs Supervision. Retrieved August 14, 2024, from https://digitalpaper.stdaily.com/http_www.kjrb.com/kjrb/html/2023 - 04/03/content_551468.htm
[36]. Zhang, X., & Ruan, C. J. (2019). The Crime Risk and Governance of Illegal Application of Artificial Intelligence. *Studies on Socialism with Chinese Characteristics*, 4, 78 - 86.
[37]. Xinhua News Agency. (2024, June 21). Create and Share, Advance Hand in Hand - President Xi Jinping's Congratulatory Letter to the 2024 World Intelligent Industry Expo Evokes Strong Resonance in the Industry. Retrieved November 8, 2024, from https://www.gov.cn/yaowen/liebiao/202406/content_6958719.htm