Properties of artificial intelligence systems in the context of their use in legal activities

Mukhtar Sadykov\textsuperscript{1}, Mohammad Ershadul Karim\textsuperscript{2}, Ane Tynyshbayeva \textsuperscript{3} and Dmitry Bakhteev\textsuperscript{4}

\textsuperscript{1} Academy of Law Enforcement Agencies under the Prosecutor General’s Office, Kazakhstan
\textsuperscript{2} Universiti Malaya, Malaysia
\textsuperscript{3} Academy of Public Administration under the President of the Republic of Kazakhstan, Kazakhstan
\textsuperscript{4} Ural State Law University, Russia
mukhtar.sadykov@gmail.com

Abstract. The study was undertaken to reflect on the values of modern artificial intelligence systems in the context of the use of such systems in legal, especially law enforcement activities. The application of artificial intelligence in the research focuses on the properties of fairness, accountability, and transparency. Fairness should exclude distortions in the operation of artificial intelligence systems, caused by the settings of scales or the specifics of the dataset collected for training the system. Accountability is seen as the property of an AI system to protect user data that is included in a dataset or processed by an AI system. Transparency, on the other hand, reflects the ability to verify the decision logic of an AI system and reverse-engineer its algorithm. This property is currently the least attainable, but it is directly related to the evaluation of the effectiveness of AI, and hence to the possibilities of integrating such systems into legal activities. This paper uses the current understanding of the capabilities of systems based on machine learning methods: convolutional artificial neural networks and transformer networks. The study reveals differences and discussions of AI perspectives in legislation and the state of legal regulation, public and academic approaches to this issue in the European Union, the USA, Canada, Singapore, China, Russia, and Kazakhstan. As a result, the study proposes a set of recommendations for banning/restricting the use of artificial intelligence and decision support systems, considering national and international legislation.

Keywords: artificial intelligence, fairness, accountability, transparency, legal regulation.

1 Introduction

The importance of generative artificial intelligence (AI) is incredibly significant as it possesses the capability to produce and generate fresh content, encompassing images, texts, and even music. Its extensive range of applications and benefits spans across diverse industries. These include the generation of creative content, customization and
recommendation systems, data expansion, virtual environments and gaming, healthcare and drug exploration, natural language processing, creative tools, and design support. These examples effectively demonstrate the relevance of generative AI in various domains. The potential of this technology to automate creative tasks, enhance decision-making processes, and improve user experiences renders it a captivating and influential innovation [1].

ChatGPT’s exciting capabilities and wide range of applications draw considerable attention and make a profound impression. Its ability to exhibit high-level intelligence and intuition in a variety of scenarios, including coding, content creation, and answering a variety of questions, clearly demonstrates the power of this artificial intelligence chat tool. Using machine learning (ML) and natural language processing (NLP) techniques, ChatGPT has already demonstrated its ability to generate valuable and original content. OpenAI co-founder Elon Musk expressed his surprise in a tweet in December 2022, noting that ChatGPT’s prowess is both impressive and somewhat alarming, indicating that we are approaching a potentially dangerous stage in the development of artificial intelligence [2].

In addition to generative artificial intelligence, it is also worth noting the importance of legal procedures for creating machine-readable law: a set of technological operations for the digitization of paper documents, and the formation of fully automated court proceedings while respecting the principles of transparency and accountability.

2 Problem statement

Although artificial intelligence technology is objectively effective. However, the processes of its development, implementation, use, and evaluation of the results are important factors to be examined as well as the features of the mentality of the legal community, legislative attitudes, and positions in the technical, and humanities, typical of individual states, and interstate legal families.

3 Research Questions

1. How do national legal systems respond to the need for normative and ethical regulation of artificial intelligence technology development?
2. What regulations govern AI technology in the European Union, USA, Canada, Singapore, China, Russia, and Kazakhstan?
3. Are the properties of fairness, accountability, and transparency taken into account in the preparation of regulations governing AI technology?
4 Purpose of Study

The purpose of this paper is to discuss the national characteristics of the legal regulation of artificial intelligence based on three key qualities - fairness, accountability, and transparency.

5 Research Methods

The main methods of research are qualitative observation, comparative legal analysis, and literature review. Given the predictive nature of the study, it also applies to the method of ethical-legal modeling.

6 Findings

6.1 Ethical issues associated with artificial intelligence

According to V. Chiao, the ethical problems associated with artificial intelligence can be divided into three categories: problems of fairness, accountability, and transparency. First, if AI uses biased and unprocessed information, there is a question about the credibility of such AI. Second, who is responsible for the negative outcomes that arise from the use of AI? Unlike humans, disputes with an algorithm can be as meaningful as disputes with a household appliance such as a refrigerator or toaster. Finally, how important is it for us to understand the inner workings of an algorithm, and what consequences can arise from not understanding the logic used by AI in making decisions [3]?

**Fairness.** It is argued that fairness should exclude distortions in the operation of artificial intelligence systems, caused by the settings of scales or the specifics of the dataset collected for training the system.

The biggest issue with fairness arises from possible bias/discrimination in AI decision-making. One of the fundamental principles supporting fairness can be one of the core principles of the European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment, developed by the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe. The principle of non-discrimination: namely, to prevent the development or exacerbation of discrimination between individuals or groups of persons [4] [5]. Tobler presents discrimination as the application of different rules/practices to comparable situations, or one rule to different situations [6].

For example, let us make two assumptions: 1) when on bail, prior arrests are a predictor (predictor) of the possibility of failing to appear in court (prosecuting authority) or committing an offense while free on bail 2) the average frequency of arrests is higher in group A than in group B. Thus, we can tentatively assume a biased attitude toward people in Group A, based on the differences in the frequency of arrests compared to Group B. If this parameter is used for bail - it would further increase the
injustice towards group A. However, it is worth noting that the presence of bias depends on how the difference between the two groups is obtained.

Compare two situations. Situation 1. Members of group A have a higher average frequency of arrests because they tend to be riskier and more aggressive. Agree that this is an objective factor where there is no room for bias and discrimination. Situation 2. Group A members are more closely scrutinized by the criminal prosecution authorities, and this is not due to objective necessity.

We agree with V. Chiao that it can be considered unfair for the predictor indicators in group A in the second situation, while in the first situation, this indicator can be regarded as objective [3].

Barocas and Selbst, specifically about big data and algorithmic decision-making, argue that discrimination can be an artifact of the data collection and analysis process itself [7].

Crawford and Schultz suggest that more specifically, even with the best of intentions, algorithmic data-driven decision-making can lead to discriminatory practices and outcomes: algorithmic decision-making procedures can replicate existing patterns of discrimination, inherit the biases of previous decision-makers, or simply reflect widespread biases that persist in society [8].

**Accountability.** It is also argued that accountability is seen as a property of an AI system to protect user data that is included in a dataset or processed by an AI system. There are two contrasting attitudes toward the utilization of artificial intelligence systems in processing user data or other restricted data:

1. If there is public trust in the concept of automated data processing, the current state of affairs will be maintained, and the possibility of manual data processing may even be eliminated in the future. In this scenario, users must receive comprehensive and easily understandable information, rather than lengthy legal texts in small print, regarding the nature of data transactions and their potential utilization, including by third parties. This approach can be summarized as “Data will be secure without human intervention.” We believe that this standpoint appears more appropriate, especially considering the evolving public mindset.

2. Automated data processing will be seen as unreliable or potentially unlawful. A shift in public perception of this nature can only occur following multiple instances of significant data breaches causing tangible harm to essential aspects of human life. [5].

An example to support the first approach is Elinar which has developed software that allows a person to be identified without the possibility of leaking personal data. For example, if a customer submits a copy of his passport to the organization, the artificial intelligence system divides the image of the document into many small parts, and the subsequent identification is made concerning these fragments, without the overall view of the document. A complete copy of the passport is not contained in the system, the human operator's access to the database may also be difficult or excluded [9].

**Transparency.** In our opinion transparency reflects the ability to verify the decision logic of an AI system and reverse-engineer its algorithm. This property is currently
the least attainable, but it is directly related to the evaluation of the effectiveness of AI, and hence to the possibilities of integrating such systems into legal activities.

Understanding the potential risks associated with the widespread use of technology is crucial. There are two primary reasons why risky situations can arise when utilizing artificial intelligence systems: errors within the system itself and instances of improper usage (targeting error). It's important to recognize that artificial intelligence systems can cause harm not only in the physical realm, such as cyber-physical systems but also in the digital domain. For instance, they can mishandle citizen applications or financial transactions.

The root cause of such risky situations lies within the technology itself, specifically the lack of complete transparency in the functioning of artificial intelligence, particularly in the form of artificial neural networks. These networks, with their adaptive learning capabilities, possess a “hidden layer” within their information processing structure. This layer engages in heuristic data processing, enabling the discovery of patterns that may elude human cognition. Consequently, the logic behind the final decisions made by artificial intelligence systems may appear obscure, rendering them unreliable or irrational.

It is important to highlight that a portion of the risks stem from the realm of technological competition. Due to intense competition, software developers find it unprofitable to disclose the algorithms used for learning and decision-making in commercial products. This further complicates the already challenging task of ensuring transparency and control over the functioning of artificial intelligence systems. Frequently, the development of an artificial intelligence system occurs within an environment of commercial secrecy. In such cases, the training algorithm and/or dataset are typically kept classified.

Furthermore, the decision to conceal the fact of the development or its specific details may also arise due to the uncertainty surrounding the outcome. This situation is more common in scientific research and experimental models where monetization is not the primary objective [5].

6.2 AI regulation (review)

European Union. Europe is leading the way in developing legislation to regulate artificial intelligence. On May 11, 2023, the leading parliamentary committees of the European Parliament approved the Artificial Intelligence Act (AI Act), drafted two years ago. The next stage will involve the adoption of the law at a plenary session. After members of the European Parliament formalize their position, the proposal will enter the final stage of the legislative process - negotiations with the EU Council and the Commission, known as the trilogue.

The AI Act serves as the main legislative framework for regulating artificial intelligence, given its potential for harm. The Act was jointly approved by the parliamentary committees on civil liberties and the internal market, garnering support from a large majority [10].

The purpose of the legislation is to guarantee human control over AI systems, ensuring their safety, transparency, traceability, non-discrimination, and environmental
friendliness. In addition, there is a desire to create a consistent definition of AI that can be adapted to different technologies, encompassing both current and future AI systems [11].

The AI Act divides AI-based computer programs into three levels of risk (unacceptable risk, high-risk applications, and applications that are not explicitly prohibited or classified as high-risk) and the degree of regulation differs based on these levels of risk.

At the same time, A. Thierer, author of the book “Evasive Entrepreneurs”, believes that innovations in AI, which appeared in the United States, will never appear in Europe by definition, since the laws simply do not allow it. He believes that the European approach to AI regulation will only strengthen the power of global IT giants because only they can maintain legal units that can bring everything in line with the norms of the AI Act [12].

**United States of America.** Currently, the United States is considered a kind of “haven” for innovation in the field of artificial intelligence, with a liberal approach to regulating this area. The U.S. is taking the first preliminary steps to establish rules for artificial intelligence tools, as the hype around generative AI and chatbots has reached its climax.

On April 11, 2023, the U.S. Department of Commerce announced that it was formally requesting public comments on how to create accountability measures for artificial intelligence and asking for help on how to advise U.S. policymakers to approach the technology [13].

The White House has proposed a “Blueprint for an AI Bill of Rights” that outlines salient principles of preventing discrimination and protecting user privacy and security. Notably, the National Institute of Standards and Technology has released an AI Risk Management framework. However, so far, Washington has taken a voluntary approach to compliance, while experts say there is a need for a more mandatory approach to AI regulation [14].

**People's Republic of China.** A draft law titled "Administrative Measures for Generative Artificial Intelligence Services," published by the Cyberspace Administration of China, states that national Internet regulatory agencies must conduct a security assessment before offering generative AI products to the public. The purpose of this law is to ensure the responsible and regulated use of generative artificial intelligence technology for its healthy development. The content generated by AI must be consistent with core socialist values and must not contain material that challenges state authority. In addition, it must not contain terrorist or extremist propaganda, encourage ethnic hatred, or any other content that could disrupt economic and social stability [15].

**Singapore.** Like many other countries, Singapore lacks a comprehensive law or regulatory framework specifically dedicated to AI. Instead, it is currently relying on existing laws, common law principles, regulatory bodies, and recently introduced national guidelines to regulate AI-related issues [16].
In April 2023, the Minister of Communications and Information said during the Parliamentary session that Singapore supports the responsible development and deployment of AI so that its benefits can be enjoyed in an environment of trust and security. To achieve this goal, an AI framework management model was introduced. Companies such as DBS, HSBC, Visa, and Microsoft have adopted this framework to address key ethical and governance issues in the implementation of AI solutions. Another tool is AI Verify, a self-test system and toolkit to demonstrate responsible AI adoption. More than 50 companies were interested in it. The Personal Data Protection Commission (PDPC) will issue advisory recommendations on the use of personal data in AI systems within a year by the Personal Data Protection Act (PDPA). The PDPC engages with industry and international partners on AI-related issues through our Advisory Council on the Ethical Use of AI and Data and the Global Partnership for Artificial Intelligence. Just as the Singapore government regulates data use, cybersecurity, misinformation, and online harm, they will continue to analyze the state of technology, the approach to regulation, and its effectiveness in maintaining trust and security in digital developments [17].

**Canada.** In June 2022, the Canadian government introduced the Artificial Intelligence and Data Act (AIDA) as part of Bill C-27, the Digital Charter Implementation Act of 2022. AIDA marks a significant milestone in implementing the Digital Charter and aims to establish trust among Canadians in the digital technologies they use daily. The legislation emphasizes the importance of ensuring safety and alignment with Canadian values in the design, development, and utilization of artificial intelligence systems.

AIDA’s concept represents the initial phase of a new regulatory framework designed to guide AI innovation in a positive direction and encourage responsible adoption of AI technologies by Canadian individuals and businesses. The government intends to develop this framework through an open and transparent process of regulation development, incorporating consultations to gather input from diverse stakeholders across Canada. The objective is to ensure that the outcomes align with Canadian values.

Given the global interconnectivity of the digital economy, the regulation of AI systems in the market necessitates international coordination. Canada actively collaborates with international partners such as the European Union, the United Kingdom, and the United States of America to harmonize approaches and ensure global protection for Canadians. Additionally, this collaboration aims to establish recognition for Canadian companies internationally by meeting robust standards [18].

**Russian Federation.** The national strategy for the development of artificial intelligence until 2030 [19] is a key document that creates a favorable environment in the Russian Federation for the creation and implementation of intelligent systems. The state, its agencies, and large corporations are investing heavily in collecting datasets and developing machine learning models in various sectors of the economy, both in the public and private sectors.
A key figure in the ethical and legal regulation of artificial intelligence development activities in Russia is the Alliance for Artificial Intelligence. The corresponding code of ethics developed by this organization stipulates only cases of civil (non-military) use, which seems to be a significant omission. In addition, from the standpoint of legal technique, the document essentially calls for collective external influence on the participants of this technology, which, by the example of many international treaties, is not optimal: in the logic of the development of any critical technology, it is enough for one violator to set a different vector of its formation. In our opinion, ethical and legal regulation should be carried out "from the bottom" and come directly from the persons involved in the development of intelligent systems. At the same time, the above drawbacks of the considered document can be easily explained by a wide range of its signatories. At present, this code provides mandatory consideration of the properties of accountability and fairness in the development of artificial intelligence systems in Russia. We believe that ignoring the criterion of transparency is due to the realization that at this stage of technological development, its observance seems unlikely.

An important aspect of the development of artificial intelligence systems in Russia is the study of the risks and limitations of implementing this technology. Thus, A. Bessonov and D. Bakhteev in their works related to the development of applied systems in law enforcement repeatedly pointed to the need for preliminary study of the aspect of human activity, which is optimized with the help of intelligent systems; the need to train law enforcement officer on the basics of technology before introducing it into legal processes; correlation of principles of forming system conclusions to evaluation criteria [21] [5].

Of the three criteria stated at the beginning of the article, fairness is perceived as the key one in Russia: artificial intelligence systems are perceived to ensure the accuracy and efficiency of legal procedures. Most surveys of the public and legal professionals show a positive attitude of citizens toward such systems. At the same time, attempts to introduce machine learning systems in judicial and expert systems face the non-fulfillment of accountability and transparency criteria, which may slow down the introduction of intelligent technologies into the Russian legal system. Accordingly, it is necessary to observe the principle: if the adoption of many decisions can be evaluated statistically, an individual decision, cannot be investigated for the correctness of its logic, in which case it should remain possible to study the situation by a qualified human specialist.

Republic of Kazakhstan. The development of artificial intelligence technology is designated as one of the main priorities of the country's development, the main factor of global progress, and one of the trends in digitalization, the development of which leads to an increase in the efficiency and speed of business, openness, and transparency of the dialogue of citizens with their states. Particular attention was paid to the development of artificial intelligence in the state program "Digital Kazakhstan", the Strategic Plan development of the Republic of Kazakhstan until 2025, which identified specific projects: the creation of a research institute for the development of artificial intelligence technologies, a national cluster of artificial intelligence and data pro-
cessing centers based on the leading university of the country the Nazarbayev University [22].

In Kazakhstan, the recognition of the necessity for the legal regulation of AI was initially expressed in 2021 within the Legal Policy Concept to 2030 [23]. This policy document identified two key factors that require the implementation of regulations for AI technologies and robotics:

1. Resolving the matter of assigning liability for any harm caused by AI and robots.
2. Addressing the challenge of determining ownership of intellectual property rights for works created with the involvement of AI.

Consequently, even though there is currently no specific legislation in place for AI regulation, Kazakhstan has identified these two primary areas that require legal frameworks to establish rules for the operation of AI [24].

In addition, Kazakhstan has enacted a law concerning personal data and its protection [25]. However, the country has not yet made explicit attempts to directly regulate the usage of AI regarding data protection. Nevertheless, considering certain provisions of the existing law, it is evident that businesses will face significant limitations in utilizing AI. For instance, if the operation of AI necessitates the use of biometric data (which falls within the scope of personal data in Kazakhstan), obtaining consent would be required for the usage of such data. Therefore, it can be inferred that the use of AI will be subject to restrictions and consent requirements based on the current legal framework governing personal data in Kazakhstan.

6.3 AI regulation (our vision)

We assume that the differences in approaches to the regulation of artificial intelligence lie in the plane of socio-cultural differences and inherently reflect the culture and psychology of the individual country i.e., the local context.

Therefore, in our opinion, we should not expect a global legal regime to regulate this area soon. Thus, we can conclude that AI governance around the world is fragmented. There are also many initiatives in this area, including codes of ethics and principles for the responsible use of AI, but these are not binding.

This problem in regulation will persist because it is rooted in two issues that underlie the governance of all new technologies, from synthetic biology to cryptocurrencies, both of which cannot be easily solved. These difficult scenarios are the pacing problem and feeding the debate on the Collingridge dilemma.

The pacing problem. The application, deployment, and diffusion of technology evolve rapidly, while laws and regulations are developed and adopted more slowly, and regulation tends to “catch up” with technology. The application of technology is universal whereas regulation is domestic or country-specific.

In addition, the development of global regulation takes an enormous amount of time and effort, and they are not always successful. This mismatch is called the pace problem.
Even worse, the pace problem is amplified by combinatorial innovations: technological and developmental capabilities that quickly and symbiotically build on each other to accelerate innovation.

**Collingridge's dilemma.** David Collingridge introduced a concept now known as the Collingridge dilemma. The dilemma is that regulating a technology in its early stages, when its potential dangers are not yet apparent, is an easy task, but becomes more difficult by the time those dangers are identified.

Early regulation is likely to be too restrictive for further development and implementation, while regulation at a more mature stage may be limited in its effectiveness and ability to prevent accidents [26].

### 7 Conclusion

As previously mentioned, the primary concern regarding fairness in AI lies in the potential bias or discrimination in decision-making. It is crucial to not only examine how the algorithm makes decisions but also to consider the data used to train the algorithm and the environment in which it operates.

We believe that the algorithm itself is just the visible part of the issue and cannot eliminate the racial biases embedded in previous stages of decision-making. These earlier decisions serve as the foundation for the variables the algorithm relies on to determine, for example.

Addressing this problem goes beyond technical decision-making. It requires a comprehensive analysis of all stages, starting from basic police patrolling and detention, all the way to the trial process. We agree with V. Chiao's description of a race-neutral process as one where no decision point introduces racial bias [3].

In this context, the analysis should address numerous questions, including: 1) What risks of racial bias exist at each step? 2) How can these risks be mitigated? 3) Should humans or artificial intelligence be responsible for decision-making at specific stages, considering that early-stage decisions should not negatively influence later-stage decisions? 4) How should conflicts between artificial intelligence and humans be resolved? Who has the final say, and based on what criteria (clear criteria should be established, free from external political, subjective, or populist influences)?

This list of questions is not exhaustive but highlights the issues that need resolution to appropriately structure the process and integrate artificial intelligence if necessary. By answering these questions, we can gain a clear understanding of how the process should be organized and, most importantly, how to maximize the benefits of using artificial intelligence. It is crucial to view artificial intelligence as an integral part of the organizational process rather than in isolation, to enhance the overall effectiveness of the process.

We also suggest utilizing the term “explicability”. This principle differs from the notion of transparency already discussed earlier in the paper. Explicability in the context of this study should be understood as the possibility of explaining the technolo-
gies of training and functioning of the system, not only to the customer but also to users and the public. Apart from this, the origin of the dataset used to train the system should also be explainable. The explicability requirement, however, must not contain a threat to the confidentiality of either trade secrets or personal data [5].

Here comes our vision of accountability as a property of an AI system to protect user data that is included in a dataset or processed by an AI system. We believe that the principle of “Data will be secure without human intervention” appears more appropriate, especially considering the evolving public mindset [5].

Above mentioned example of Elinar’s software allows a person to be identified without the possibility of leaking personal data. A complete copy of the passport is not contained in the system, the human operator's access to the database may also be difficult or excluded [9].

As a promising model of legal regulation, it is also worth mentioning the model of “ethical switch” [27]. Scaling this model from autonomous cars to all areas of AI use, we can suggest the following: the state offers several alternative options for regulating the development, implementation, and use of AI systems for the private sector of the economy. Easing controls would then correspond to increasing liability when incidents occur, and vice versa.

As of 2019, machine learning dominates AI techniques (1/3 of all inventions or 134,777 patent documents). Applications are also growing at a pace of 28% per annum. That’s why our recommendations are concentrated more towards this technique [28].

Some potential solutions to address the challenges of the pace problem and the Collingridge dilemma encompass anticipatory governance, “soft laws,” and the utilization of regulatory sandboxes [26].

Anticipatory governance involves using foresight about future events to shape current policies and practices. Regular and meaningful engagement with stakeholders and flexible governance mechanisms contribute to improved anticipation.

“Soft laws” refer to voluntary guidelines, industry-established standards, and consensus principles that often involve limited regulatory involvement. While soft laws may lack legal enforceability, they establish clear boundaries and can complement existing regulations.

A regulatory sandbox serves as a controlled environment where innovators can experiment with new products or services under regulatory supervision. This process allows regulators to gain insights into the technology, its intended application, and the choices it presents to stakeholders.

Implementing these strategies could help address the pacing problem and the Collingridge dilemma, providing regulators with some level of control and predictability regarding AI. However, it is challenging to determine whether these approaches will serve as the optimal solution at this stage.
Properties of artificial intelligence systems in the context

References


20. Kodeks jetiki v sfere iskusstvennogo intellekta // Al'jan v sfere iskusstvennogo intellekta, https://a-ai.ru/wp-content/uploads/2021/10/%D0%9A%D0%BE%D0%B4%D0%B5%D0%BA%D1%81%D1%8D%D1%82%D0%B8%D0%BA%D0%B8_%D0%B2_%D1%81%D1%84%D0%B5%D1%80%D0%B5_%D0%98%D0%98_%D1%84%D0%B8%D0%BD%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9.pdf, last accessed 2023/05/26.


