Information for citation:

Gulyamov S. S. Mechanisms for Building Trust in AI-Assisted Justice Systems Through Cognitive, Emotional and Socio-Psychological Factors *Vestnik Permskogo universiteta. Juridicheskie nauki* – Perm University Herald. Juridical Sciences. 2025. Issue 3(69). Pp. 482–491. (In Russ.). DOI: 10.17072/1995-4190-2025-69-482-491.

Gulyamov S. S. Mechanisms for Building Trust in AI-Assisted Justice Systems Through Cognitive, Emotional and Socio-Psychological Factors // Вестник Пермского университета. Юридические науки. 2025. Вып. 3(69). С. 482–491. DOI: 10.17072/1995-4190-2025-69-482-491.

UDC 347
DOI: 10.17072/1995-4190-2025-69-482-491

Mechanisms for Building Trust in AI-Assisted Justice Systems Through Cognitive, Emotional and Socio-Psychological Factors

S. S. Gulyamov
Tashkent State University of Law
E-mail: said.gulyamov1976@gmail.com

Received 07 Apr 2025

Introduction: *artificial intelligence (AI) is progressively gaining traction within legal and judicial spheres, which necessitates a re-examination of concepts such as transparency, fairness, reliability, and legitimacy. The implementation of explainable artificial intelligence (XAI) is a key element in ensuring trust in such systems, though 'explainability' does not in itself guarantee that people will begin to trust these systems.* ***Purpose****: the study aims to provide a detailed analysis and systematization of available scientific information on psychological mechanisms underlying the formation of trust in explainable artificial intelligence systems applied in the field of justice, with the cognitive, emotional, and socio-psychological aspects considered.* ***Methods****: the study employed methods of analysis and synthesis to explore the complex system of the formation of trust in explainable artificial intelligence in the field of justice. An analysis of a wide range of open sources, including scientific articles, legal documents and research reports, made it possible to categorize facts regarding the perception of computerized judicial decisions and the level of trust in them.* ***Results****: the research has established that trust in AI applied in the field of justice develops at the intersection of three key aspects: the understanding of the fundamentals of AI operations, emotional acceptance of the technology, and its adherence to social norms. The author explains the significance of transparency and explainability of AI decisions, of the ability of AI systems to interact with users on an emotional level, and of the role of public opinion and expert examination in trust development. The study has found correlations between perceived accuracy of AI systems and trust levels, elucidated the importance of empathetic characteristics in AI interfaces and the influence of group dynamics on trust formation.* ***Conclusions****: the findings demonstrate the need for a comprehensive approach to AI implementation in the legal sector that would consider both psychological and technological factors. The paper provides practical recommendations for increasing trust in AI in justice, including the development of training programs, standardized explainability schemes, and transparent audit mechanisms. Recommended areas of future research include studies into cross-cultural variations in attitudes to AI, long-term strategic consequences and ramifications of justice automation, and the development of methods for assessing the effectiveness of human-AI collaboration in court.*

Keywords: artificial intelligence; justice; trust; explainability; cognitive factors; emotional perception; socio-psychological aspects

© Gulyamov S. S., 2025

This work is licensed under CC BY 4.0. To view a copy of the license, visit https://creativecommons.org/licenses/by/4.0/

Механизмы формирования доверия к системам правосудия, использующим искусственный интеллект: когнитивные, эмоциональные и социально-психологические факторы

С. С. Гулямов
Ташкентский государственный юридический университет
E-mail: said.gulyamov1976@gmail.com

Статья поступила в редакцию 07.04.2025

Введение: *искусственный интеллект (ИИ; англ.: AI) активно внедряется в правовую и судебную сферы, что требует переосмысления концепций прозрачности, справедливости, надежности и легитимности. Внедрение объяснимого искусственного интеллекта (XAI) является ключевым элементом для обеспечения доверия к таким системам, однако сама по себе «объяснимость» не гарантирует, что люди начнут доверять системам.* ***Цель:*** *провести детальный анализ и систематизировать доступную научную информацию о психологических механизмах формирования доверия к системам объяснимого искусственного интеллекта в сфере правосудия с учетом когнитивных, эмоциональных и социально-психологических аспектов.* ***Методы:*** *методы анализа и синтеза были использованы для изучения процессов формирования доверия к объяснимому искусственному интеллекту в правосудии. Был проведен анализ широкого спектра открытых источников, включая научные статьи, юридические документы и отчеты исследований, что позволило категоризировать факты, касающиеся восприятия компьютеризированных судебных решений и уровня доверия к ним.* ***Результаты:*** *исследование показало, что доверие к ИИ в правосудии формируется на пересечении трех ключевых аспектов: понимание основ работы ИИ, эмоциональное принятие технологии и соответствие технологии социальным нормам. Статья объясняет значимость прозрачности и объяснимости решений ИИ, способности систем эмоционально взаимодействовать с пользователями, а также роли общественного мнения и экспертной оценки в развитии доверия. Обнаружены корреляции между воспринимаемой точностью систем ИИ и уровнем доверия к ним, обоснована важность эмпатических характеристик в интерфейсах ИИ и установлено влияние групповой динамики на формирование доверия.* ***Выводы:*** *результаты исследования демонстрируют необходимость комплексного подхода к внедрению ИИ в правовой сфере с учетом как психологических, так и технологических факторов. Сформулированы практические рекомендации по повышению доверия к ИИ в сфере юстиции, включая разработку программ обучения, стандартизированных схем объяснимости и прозрачных механизмов аудита. Перспективные направления дальнейших исследований включают изучение межкультурных различий в формировании того или иного отношения к ИИ, анализ долгосрочных стратегических последствий автоматизации правосудия и разработку методов оценки эффективности сотрудничества человека и ИИ в суде.*

Ключевые слова: искусственный интеллект; правосудие; доверие; объяснимость; когнитивные факторы; эмоциональное восприятие; социально-психологические аспекты

Introduction

Artificial intelligence is progressively gaining traction within legal and judicial spheres and constitutes a paradigm shift in the understanding of justice, which necessitates a re-examination and reconstruction of society's perception of issues such as transparency, fairness, reliability, and legitimacy[1]. In 2019 the United Nations Office on Drugs and Crime (UNODC) released a report highlighting the potential of artificial intelligence to achieve a better understanding of the nature and mechanisms of judicial processes so as to preclude algorithmic bias, ensure data privacy, and avoid erosion of human judgment in legal decision-making[2]. Human understanding of the reasoning and operation of such systems is ensured through the concept of explainable AI (XAI).

[1] European Commission. White Paper on Artificial Intelligence: A European approach to excellence and trust, 2020. Available at: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.
[2] UNODC. Artificial Intelligence and Robotics for Law Enforcement, 2019. Available at: https://www.unicri.org/artificial-intelligence-and-robotics-law-enforcement.

According to Article 13 of the Artificial Intelligence Act[1], high-risk systems must be developed with consideration of all cybersecurity aspects, under appropriate monitoring by human specialists, and specific mechanisms are to be implemented in order to effectively and professionally review the output of such systems. Meanwhile, even the 'explainability' feature of artificial intelligence is no guarantee that humans will start trusting AI systems, these based on neural networks, machine learning, deep learning, reinforcement learning, and other technologies covered by the general term 'artificial intelligence'. The objective of the study is to carry out a detailed analysis and systematization of available scientific information on psychological mechanisms underlying the formation of trust in explainable AI systems in the area of justice, while taking into account cognitive, emotional, and socio-psychological aspects. Cognitive, emotional, and socio-psychological processes play the biggest role in how individuals perceive computerized legal systems and judge whether they are objective, just, and proficient. The present paper is based on what was established in the 2018 European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems[2]. The Charter calls for adherence to basic principles such as respect for human rights, non-discrimination, security, transparency, and maintaining user control, particularly in sensitive law enforcement areas. Examination of these principles from a perspective of psychological trust mechanisms allows for a greater comprehension of AI implementation issues in the judicial field. An in-depth study of the determinants of trust in AI in the legal field is supposed to make a rich contribution to the debate on artificial intelligence regulation in the legal field. Providing a typology of the determinants offers a foundation for formulating efficient strategies and policies for building public trust in AI-enabled justice systems. This is the way to maintain basic values of a democratic state and ensure the rule of law in the implementation of new technologies in the judiciary.

Methodology

In the present research, the author employed the methods of analysis and synthesis of psychological processes and investigation into the complex process of the formation of trust in explainable artificial intelligence when applied in the justice system. The research methodology underlying this study was guided by international legal documents such as the Recommendation of the Council on Artificial Intelligence of 2019[3], which point to stakeholder collaboration and evidence-based evaluation when operating, managing, and controlling artificial intelligence. An examination of a wide variety of open sources, such as scientific articles, legal documents, and survey reports [2; 4; 37; 32; 10; 3, 17; 12; 9; 35; 33; 5; 6; 34; 7], as well as other types of research works listed in the references, made it possible to categorize facts concerning perception and the level of trust in computerized judicial rulings. A comprehensive analysis of diverse software tools used in legal practice [8] and review of relevant literature [10; 3; 14; 27; 23] helped to define major patterns and priority areas of research. This allowed for insight into the context of interaction of cognitive, emotional, and socio-psychological factors in the process of development of mechanisms of trust in AI systems in the area of jurisprudence.

Understanding the Principles of AI

An analysis of the level of trust in artificial intelligence systems in the justice sector with cognitive factors taken into account reveals a close relationship between the trust level and the degree of public understanding of the principles of AI operation. This understanding includes a set of theoretical and practical knowledge, as well as awareness of the potential and real application of AI systems in the context of law and legal proceedings. From a theoretical point of view, this includes concepts such as algorithms related to machine learning, neural systems, decision trees. From a practical point of view, this includes legal analytics, decision support systems, case management, etc. As various studies show, there is a positive correlation between the results that artificial intelligence systems produce and the level of trust in them [1; 29; 11].

In addition, according to studies, information-literate people demonstrate a 30% higher level of trust compared to those who do not have this level of knowledge or are in a state of digital nihilism[4]. 'Digital nihilists' can be described as people who tend to exaggerate fears and risks associated with such phenomena as handling artificial intelligence, using digital services, banks, transactions, cryptocurrencies, and other benefits of our century, while those fears are unfounded and not supported by practice.

The level of knowledge and understanding of the functioning and intended purpose of AI systems can vary significantly. The nature and degree of the correlation between such literacy and trust in AI depend on the specifics of a particular system and the context of its application in the legal sphere.

[1] European Commission. Proposal for a Regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act), 2021. Available at: https://artificialintelligenceact.eu/.

[2] Council of Europe. European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment, 2018. Available at: https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment.

[3] OECD. Recommendation of the Council on Artificial Intelligence, 2019. Available at: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

[4] Berryhill J., Heang K., Clogher R., McBride K. (2019). *Hello, World: Artificial Intelligence and Its Use in the Public Sector.* Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/11/hello-world_7734f030/726fd39d-en.pdf

A good example here is Estonia, where population demonstrates a high level of digital literacy and where public trust in digital technologies, according to mass surveys, is around 80%[1]. This case convincingly shows that for building trust in justice systems that are based on electronic legal proceedings, it is, first of all, necessary to take measures aimed at improving digital literacy in the population so as to leave no room for prejudices, distortions, and fears.

Assessing the Accuracy and Reliability of an AI system

AI systems can operate in different ways, and the public perception of how accurately and reliably a particular AI-assisted system operates ultimately influences public trust in the judicial system as a whole. The factors behind positive perception are algorithmic accuracy, lack of bias, risk mitigation, resilience to cyberattacks, and adherence to program protocols when ensuring security. An empirical analysis shows that there is a very strong correlation between how effective a system appears to be and how much it is trusted. A negative example here is a system called COMPAS [17], which assesses the potential risk of recidivism. Although this program has been implemented in many organizations, public trust was low because many of the results were very inaccurate and unfair, violating the principles of ethics, mutual trust, non-discrimination, etc.

According to other studies, systems that achieve the trust level of at least 50% work very accurately, and all reports and statistics indicate that their violations and failures in operation reach almost 0% [12]. This suggests that information that is transmitted to interested parties and disseminated to the masses must be provided on terms of high transparency and accuracy, since there is a strict correlation between the accuracy of an AI system's operation and the level of trust in it.

The Perception of Explainability of AI Decisions

Explainable AI is a versatile solution to build trust in law enforcement and judicial processes. This concept enables the development of frameworks for analyzing AI outputs that support human decision-making and interpretation in legal contexts. In addition, research shows that there is a strong correlation between how explainable and understandable an AI is and the level of trust that users have in it [9].

AI systems applied in legal contexts must incorporate mechanisms that provide compelling evidence and justification for their decisions. This conclusion suggests itself after one of the studies where participants were provided with two systems functioning on different technologies: one worked on the basis of complex generative neural networks, and the other – on the basis of a simpler interpretable and understandable artificial intelligence based on decision trees. As this study showed, among the participants who used AI systems for legal applications, the simpler system, which produced more understandable results, while maintaining the same level of accuracy and fairness, inspired 60% more trust than the technically complex but less comprehensible system [35].

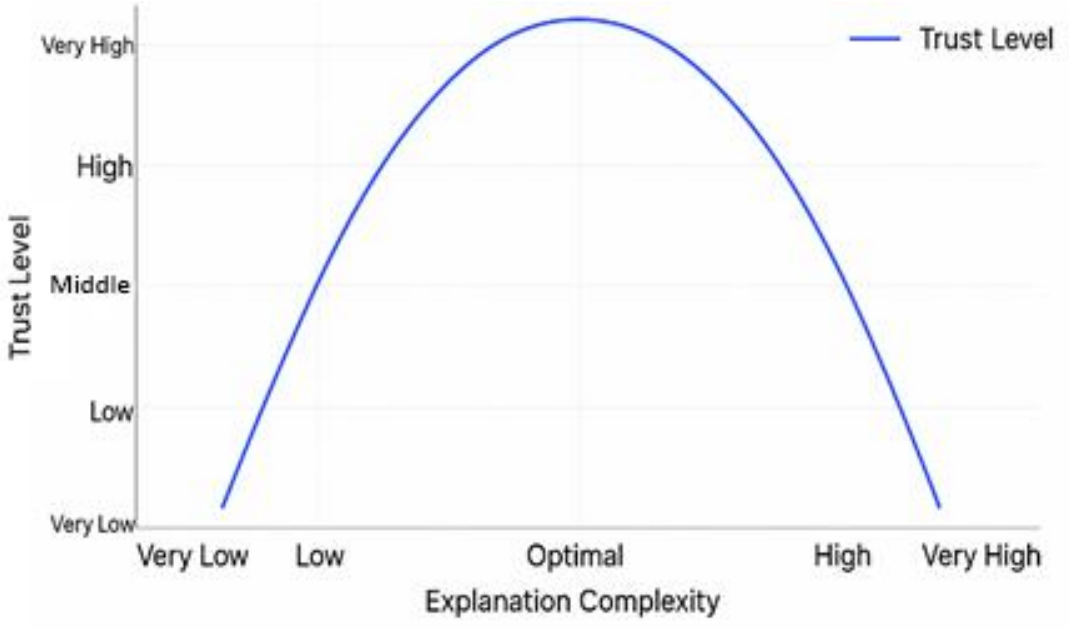


Fig. 1. The relationship between AI explainability and trust levels

[1] e-Justice factsheet, 2022. *e-Estonia* [website]. Available at: https://e-estonia.com/wp-content/uploads/factsheet_e-justice.pdf.

As shown in Fig. 1, there is a direct relationship between the level of public trust in AI systems and the level of complexity of the system, including its interpretability, up to the point where the system becomes too complex to understand, after which the relationship becomes inverse and the level of trust begins to decrease accordingly.

There is another system, called HART (Harm Risk Assessment Tool), which helps to understand whether the operation of an AI system is sufficiently clear for humans. [33]. It was developed by Dutch scientists and shows what approach should be taken to the transparency and interpretability of artificial intelligence results, and also selects which developers, working in a particular field, especially those related to the judicial and legal field, should be involved.

However, in some cases, there is no strict correlation between perceived competence, interpretability, explainability, on the one part, and trust level, on the other part: as some research findings show, if a system is not sufficiently complex, while the task being solved requires a high level of detail, the solutions that are used in simpler artificial intelligence systems will not be suitable [5]. Therefore, it is necessary to find a balance that would satisfy both the developers and the users and would promote user trust in artificial intelligence, which is possible to achieve when users understand the principles on which the AI system is built and when it is guaranteed that its results will not violate the ethical and humanistic principles. Therefore, it is very important to establish approaches to designing systems that would allow interaction between artificial intelligence and people, and to develop flexible interpretation mechanisms that would produce an analysis adapted to the complexity of the system and to user needs.

Emotional Aspects of Trust in AI in the Legal Sphere

Fear of Technology and Automation

The level of trust in artificial intelligence systems related to the justice system is strongly correlated with the concept of technophobia, or fear of using technical means. It is a psychological phenomenon manifested in cognitive-affective reactions to threats perceived by people from innovative technologies. As various studies show [6], there is an inverse relationship between how severe a person's technophobia is and what is the level of his/her trust in certain solutions that are applied with the use of artificial intelligence technologies. One of the striking examples to be given is the Austrian pilot program [34] that helps analyze documents in administrative courts using artificial intelligence. At first, there was strong resistance, and people did not trust this project as they were afraid that artificial intelligence systems would not be able to cope with those legal interpretations that they considered subtle, sensitive, and complex. However, thanks to the advertising campaigns conducted by the organization and to public involvement, it was possible to achieve impressive results: the level of citizens' trust increased by more than 40% over 6 months of the campaign [7]. This example clearly shows that the emotional resistance at the first stages of the introduction of an AI system, including a system that shows excellent accurate results, is a common phenomenon. Thus, the emotional aspects related to the acceptance of such technologies must be worked through gradually and with due understanding.

Empathy and the Perception of the 'Humanity' of AI Systems

AI systems that integrate anthropomorphic characteristics into their machine learning mechanism may be designed more effectively under the Computers Are Social Actors (CASA) paradigm. This concept holds that people tend to trust systems that demonstrate socially acceptable behavior and empathy [24]. Perceived empathy levels are highly correlated with levels of trust in legal applications. An example is an AI-powered online dispute resolution system in British Columbia [30]. This system uses sophisticated natural language processing specially designed to consider emotional nuances and criteria, which has been shown in research to increase user satisfaction by over 55% compared to conventional online systems [38]. The software showed to be more stable, and the level of perception of emotions and the nature of reactions of certain users played a significant role. This case makes it clear that integrating elements of emotional intelligence into an AI system with contextual awareness increases the level of trust in its application in the law enforcement sector. Therefore, human-centered AI that can navigate the incredibly complex landscape of basic emotions in such a challenging field as legal disputes is becoming increasingly in demand.

Emotional Attitude Toward Fairness of AI Decisions

Fairness is often subjective. Meanwhile, from the perspective of trust formation it is essential that AI systems provide decisions devoid of bias and discrimination. Algorithmic fairness, which is the result of applying a program that ensures technical accuracy and analyzes the results taking into account the 'past direction', i.e., various past cases in the relevant area, as well as cultural traditions, social ideas, and other such aspects, should be the focus of attention in such an area as justice.

Practice provides many questionable examples. For instance, facial recognition systems used in the UK, though being highly effective, showed discrimination against ethnic minorities, which suggests that biases and discrimination were perpetuated. This case tells us that such factors should become a focus of unprecedented attention, and that, before implementing a product in practice, a series of tests is to be carried out, with developers, participants, and testers coming to a conclusion that the system works without such gaps. As another study has shown, transparency with regard to how many potential biases were eliminated during beta testing, as

well as what data was provided to train artificial intelligence, improves the accuracy of its results up to 70% [15].

All this suggests that it is necessary to build holistic approaches to systems based on the principles of fairness and transparency taking into account the factors that influence the level of public trust in them (Table 1), and also to develop algorithms that can be interpreted and explained to society [39].

Table 1

Factors of trust in AI when applied in justice

Factor	Impact on trust level
Perceived empathy	Significant increase
Transparency of algorithms	Significant increase
Expert approval	Noticeable increase
User training	Strong increase over time
Interpretability and understandability of AI	Significant increase
Technophobia	Significant reduction
Pre-testing in other areas	Significant increase
AI as a complement to human labor	Moderate increase
Algorithmic fairness	Significant increase
Social norms and expectations	Moderate to strong increase
Group dynamics	Moderate to strong increase

Socio-Psychological Factors of Trust in AI in Justice

The Influence of Social Norms and Expectations

Social expectations and norms are an important factor in ensuring trust in such a complex phenomenon as artificial intelligence when applied in law enforcement. Social norms and expectations influence how people perceive and adopt technologies. There is a strong correlation between how society accepts artificial intelligence based on certain technologies in other areas and how society will treat artificial intelligence systems operating on similar technologies in law enforcement. For example, this has been proven in practice in Singapore [20], where relevant programs used intelligent case management systems that had previously been tested in other areas, and these programs gained a more than 65% higher level of trust than a system that had not previously been tested in practice [31].

This suggests that before introducing AI systems in areas such as litigation, it seems advisable to test them in less sensitive areas, such as law enforcement. Artificial intelligence should exist in close connection with legal processes, while, as studies show, the level of trust in AI systems increases by 40% when the system of neural networks and other technologies is an addition to human work, and not a replacement for it [19].

The Role of Authorities and Expert Opinion

An analysis of the opinions of authoritative experts who have a significant influence on the public perception of AI systems points to the significant role of epistemic trust in various contexts of AI application in the legal field. In addition, as showed an analysis of the literature, the endorsement by an expert council and authoritative figures in the field of artificial intelligence has a very strong effect on strengthening the level of public trust in explainable AI systems. For example, PROMETEA, which is an Argentinian program implemented in courts and prosecutors' offices, was first treated with great skepticism, but was highly praised by senior prosecutors and experts in the ethics of artificial intelligence [8]. Public trust in this system increased by more than 50% after a series of open seminars and lectures were held under the guidance of expert councils, explaining what the tasks of this system are, what tasks it can perform, what security measures are provided for its functioning [30].

Research also shows that criticism and negative feedback have a detrimental effect on trust levels, suggesting that a balanced approach and fair judgment of AI systems is needed.

Group Dynamics in the Perception of AI systems

Group use of certain AI systems is a phenomenon suggesting that the polarization of group tendencies and identity in a group can shape certain judgments and create a certain attitude and level of trust in explainable artificial intelligence in the field of justice (Fig. 2).

For example, the above-mentioned HART system was first accepted with a negative attitude, but after a course on the use of this system, where all its functions, shortcomings, advantages, and disadvantages were elucidated, the level of trust among officers increased by 60% over about 1 year [25]. This suggests that there are certain group dynamics, and the level of trust in a system can decrease or, conversely, increase, depending on how well-developed the AI system is and whether the users have managed to master the technology. In addition, advertising, mass media, and social media have a considerable impact on trust since collective perception is a highly important factor in today's world. For example, there are various communication strategies and marketing moves, with the involvement of various parties, persons, authoritative sources, and other mechanisms, that help to impact certain groups in a certain way so as to form a certain attitude and level of trust.

Fig. 2 clearly shows that there are interrelated factors influencing group dynamics in the perception of modern technologies, including AI, that determine and shape overall trust in AI systems.

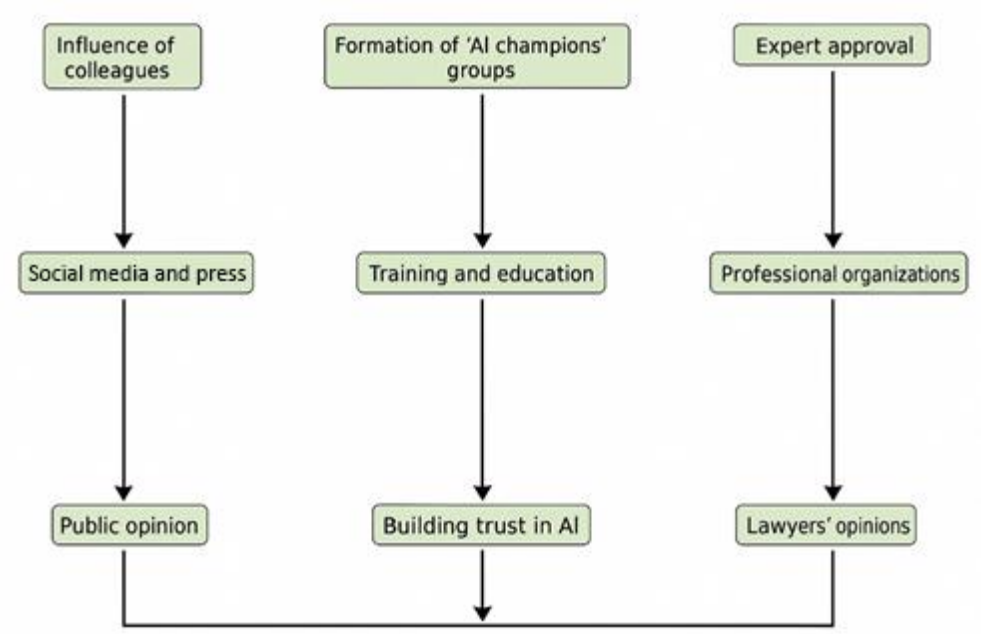


Fig. 2. Group dynamics in the formation of trust in AI

Transformation of Legal Consciousness under the Influence of AI

Changing Ideas about Justice

The modern world is undergoing a technological-legal transformation, which is manifested in the formation of an interaction between human judgment and decision-making with the use of algorithms. A good example is the e-justice system in Estonia, which has undergone a multi-stage development and implementation process. According to surveys, more than 70% of Estonian citizens rate this system as fair. A public opinion analysis also shows that the majority of respondents consider traditional justice systems without the use of artificial intelligence to be outdated and ineffective as the only instrument for administering justice [22]. Therefore, the level of digital trust is very high in this country. However, the population of many other countries, even technologically advanced ones, demonstrates a commitment to the idea that only non-automated systems represent justice. All this refers us to the concept of digital legal acculturation, which suggests that the formation of legal consciousness in the field of digital technologies affects the level of trust in general and shapes the attitude toward the effectiveness of certain systems that can complement justice with artificial intelligence in a given country.

The Evolution of the Concept of Legitimacy of Judicial Decisions

The legitimacy of judicial decisions, as a concept, is also undergoing a number of changes. There has emerged a concept of algorithmic legitimacy, which, in turn, reflects the emergence of a hybrid concept of legal authority, this combining human judgments formed by specialists and knowledge that is provided by artificial intelligence and supplements human knowledge.

For instance, in France there has been implemented a case management system called Predictice [21]. It is an AI-based platform that helps analyze historical court decisions to predict and create predictive models for various types of legal research. Initially, this system was treated by the legal community with some skepticism. However, after an information campaign was run to explain the benefits of the system and its further improvement, the level of trust has increased significantly. According to research, 65% of the surveyed lawyers noted that AI technologies effectively contribute to legal research, both from a scientific and from a practical point of view.

This example shows an expansion of the concept of legitimacy: artificial intelligence represents a combination of human and machine interaction, while functioning on objective data, and all this forms a new era based on the construction of objectivity and trusting relationships between machines and humans in a hybrid form.

Practical Recommendations for Increasing Trust in AI in Justice

Increasing trust in justice systems based on artificial intelligence involves a number of theoretical and practical recommendations that should be based on strategies taking into account various practical aspects [28]. When implementing comprehensive training programs devoted to the use of artificial intelligence in the legal field, it is essential that the principles of AI operation and its limitations, as well as the approach of explainable AI itself, be explained taking into account the principles of transparency and fairness [13; 36].

Standardized explainability schemes suggest that AI applications in the judicial field should facilitate understanding and transparency. In addition, a design that incorporates emotional intelligence, empathy, and

various other aspects in the interface of AI will also help build trust and enhance the ability of stakeholders to participate in the development of the system [11]. Furthermore, transparent audit mechanisms, ongoing reviews, international cooperation can help strengthen the creation of diverse points of view on improving user acceptance. It is also important to emphasize the need to follow pro-regulation trends worldwide, to adhere to global standards and codes that promote differentiation and consistency without violating these principles embedded in the system of artificial intelligence in the field of justice. These recommendations are aimed at creating a reliable system that could increase the level of trust in systems based on artificial intelligence technologies.

Discussion

The psychological mechanism of the formation of trust in AI systems within the legal field involves a sophisticated interaction of cognitive, emotional, and socio-psychological factors [18; 26]. Research findings show that public confidence in the use of AI systems in the judiciary depends not only on the technical efficiency of such systems but also on whether they meet social expectations and moral standards. Another central component is the explainability of AI, which is when users can see the rationale behind the decision-making. Such users' understanding and awareness subsequently facilitate the establishment of trust since users can make rational judgements concerning the system's competence. Research shows, however, that complete trust is usually unattainable due to the inherent limitations of purely rational assessment of AI systems (people cannot fully rationally evaluate all aspects of operation of complex AI systems, which prevents complete trust formation) [16]. Emotional factors, such as the ability of an AI system to empathize emotionally, are essential in establishing emotional trust. This makes it necessary to develop AI systems that can not only effectively address legal matters but also empathize with humans.

Particular attention should be given to socio-psychological processes behind the formation of trust in AI justice. Studies find social norms, group behavior, and influential voices to significantly influence the development of trust in machine-based mechanisms of justice.

This calls for an overarching approach to introducing AI in the legal system, considering not only technological but also social contexts. One of the important areas for future research is investigation into the long-term impact of AI implementation in justice on the attitudes developing in the general public with regard to fairness and legitimacy of the judiciary. Another potential research area is a deeper examination of cultural variations in the perception of AI in the legal field, which can be useful for the development of more universal and adaptive systems. Particular attention should be given to the question of how the combination of the human element and automation in judicial decision-making influences public confidence in the justice system as a whole. Thus, there is a clear need for additional interdisciplinary research integrating expertise in the areas of psychology, law, and information technology.

Conclusion

The research conducted is a comprehensive study into psychological mechanisms of building trust in explainable artificial intelligence within the legal system and legal process. Based on systematization and critical analysis of available scientific information, the paper reveals the main cognitive, emotional, and socio-psychological determinants of public perception of AI systems in the legal context.

The research findings indicate the need for a multidimensional strategy toward the AI development and deployment in the legal field. The main components of this strategy should include transparency and explainability of AI systems, consideration of emotional parameters of interaction with users, and adherence to social norms and values. The paper demonstrates that trust in AI in justice is established at the confluence of rational comprehension, emotional acceptance, and social approval.

It is necessary to develop general approaches for evaluating and enhancing the performance of human-AI collaboration within the legal process. This should involve the formation of mechanisms for increasing the degree of trust in hybrid decision-making systems and the development of training programs for legal professionals and the general public on the best modes of interaction with AI systems in the legal environment.

The future also looks promising for research on the ethics of the use of AI in justice, including the topics of algorithmic transparency, personal data protection, and algorithmic discrimination prevention. This field of research is highly correlated with the development of a normative set of rules governing AI application in the judiciary.

Another major field of future research is the examination of psychological mechanisms of society's adaptation to the growing involvement of AI in the legal field. This involves research on the transformation of legal consciousness of citizens, on their expectations from the judicial system, and on their readiness to trust technological solutions in the matters of justice.

All of these areas of research are important for ensuring harmonized development of AI in line with ethical standards and moral values. Such studies are expected to facilitate the formation of durable trust in AI systems in the field of justice through providing a foundation for the successful introduction of technologies in the legal sphere and safeguarding fundamental principles of justice and the rule of law in the digital era.

References

1. Afroogh S., Akbari A., Malone E., Kargar M., Alambeigi H. Trust in AI: Progress, Challenges, and Future Directions. *Humanities and Social Sciences Communications*. 2024. Vol. 11. Pp. 1–31. DOI: 10.1057/s41599-024-04044-8. (In Eng.).

2. Araujo T., Helberger N., Kruikemeier S., de Vreese C. H. In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence. *AI & Society*. 2020. Vol. 35. Issue 3. Pp. 611–623. (In Eng.).

3. Brkan M. Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond. *International Journal of Law and Information Technology*. 2019. Vol. 27. Issue 2. Pp. 91–121. (In Eng.).

4. Buocz T. J. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary. *Retskraft–Copenhagen Journal of Legal Studies*. 2018. Vol. 2. Issue 1. Pp. 41–59. (In Eng.).

5. Chamola V., Hassija V., Sulthana A., Ghosh D., Dhingra D., Sikdar B. A Review of Trustworthy and Explainable Artificial Intelligence (XAI). *IEEE Access*. 2023. Vol. 11. Pp. 78994–79015. DOI: 10.1109/ACCESS.2023.3294569. (In Eng.).

6. Cilliers E. J. The Challenge of Teaching Generation Z. *PEOPLE: International Journal of Social Sciences*. 2017. Vol. 3. Issue 1. Pp. 188–198. (In Eng.).

7. Cui Y. *Artificial Intelligence and Judicial Modernization*. Springer Singapore, 2020. 224 p. DOI: 10.1007/978-981-32-9880-4. (In Eng.).

8. Estevez E., Fillottrani P., Linares-Lejarraga S. *PROMETEA: Transformando la administración de justicia con herramientas de inteligencia artificial*. DOI: 10.18235/0002378. (In Span.).

9. Ferrario A., Loi M. How Explainability Contributes to Trust in AI. *SSRN Electronic Journal*. 2022. Pp. 1–24. DOI: 10.2139/ssrn.4020557. (In Eng.).

10. Fine A., Berthelot E. R., Marsh S. Public Perceptions of Judges' Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice. *Behavioral Sciences.* 2025. Vol. 15. Issue 4. Pp. 476–491. DOI: 10.3390/bs15040476. (In Eng.).

11. Gillespie N., Lockey S., Curtis C., Pool J., Akbari A. *Trust in Artificial Intelligence: A Global Study*. The University of Queensland and KPMG Australia 2023. 82 p. DOI: 10.14264/00d3c94. (In Eng.).

12. Rittenberg B., Holland C., Barnhart G., Gaudreau S., Neyedli H. Trust with Increasing and Decreasing Reliability. *Human Factors*. 2024. Vol. 66. Issue 12. Pp. 2569–2589. DOI: 10.1177/00187208241228636. (In Eng.).

13. Gulyamov S. S., Rodionov A. A. The Use of Cyber Hygiene as an Effective Psychological Measure for the Prevention of Cyber Addiction. *Psychology and Law*. 2024. Vol. 14. Issue 2. Pp. 77–91. DOI: 10.17759/psylaw.2024140206. (In Eng.).

14. Hacker P., Krestel R., Grundmann S., Naumann F. Explainable AI under Contract and Tort Law: Legal Incentives and Technical Challenges. *Artificial Intelligence and Law.* 2020. Vol. 28. Issue 4. Pp. 415–439. (In Eng.).

15. Haresamudram K., Larsson S., Heintz F. Three Levels of AI Transparency. *Computer*. 2023. Vol. 56. Issue 2. Pp. 93–100. DOI: 10.1109/MC.2022.3213181. (In Eng.).

16. Khokhlova N. I., Plekhanova N. P. *Psikhologicheskie faktory, obuslavlivayushchie doverie potrebiteley k yuridicheskoy kompanii* [Psychological Factors Contributing to Clients' Trust for a Law Company]. *Psikhologiya i pravo* – Psychology and Law. 2022. Vol. 12. Issue 1. Pp. 196–208. DOI: 10.17759/psylaw.2022120115. (In Russ.).

17. Larson J., Mattu S., Kirchner L., Angwin J. How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica*. 2016. Vol. 23. Pp. 1–16. (In Eng.).

18. Leshkevich T. G. *Paradoks doveriya k iskusstvennomu intellektu i ego obosnovanie* [The Paradox of Trust in Artificial Intelligence and Its Rationale]. *Filosofiya nauki i tekhniki* – Philosophy of Science and Technology. 2023. Vol. 28. Issue 1. Pp. 34–47. DOI: 10.21146/2413-9084-2023-28-1-34-47. (In Russ.).

19. Loh E. Medicine and the Rise of the Robots: A Qualitative Review of Recent Advances of Artificial Intelligence in Health. *BMJ Leader*. 2018. Vol. 2. Issue 2. Pp. 59–63. (In Eng.).

20. Soh J. T. H., Lim H. K., Chai I. E. Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments. *Proceedings of the Natural Legal Language Processing Workshop*. 2019. Pp. 67–77. DOI: 10.18653/v1/W19-2208. (In Eng.).

21. *Predictice*. Available at: https://www.legal-technologyhub.com/vendors/predictice/. (In Eng.).

22. *Liberties Rule of Law Report 2024: Estonia*. Civil Liberties Union for Europe, 2024. Available at: https://dq4n3btxmr8c9.cloudfront.net/files/jof46t/ESTONIA_Liberties_RuleOfLaw_Report_2024.pdf. (In Eng.).

23. Morison J., Harkens A. Re-Engineering Justice? Robot Judges, Computerized Courts and (Semi) Automated Legal Decision-Making. *Legal Studies*. 2019. Vol. 39. Issue 4. Pp. 618–635. (In Eng.).

24. Nass C., Yen C. *The Man Who Lied to His Laptop: What We Can Learn about Ourselves from Our Machines*. Penguin, 2010. 240 p. (In Eng.).

25. Oswald M., Grace J., Urwin S., Barnes G. C. Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality. *Information & Communications Technology Law*. 2018. Vol. 27. Issue 2. Pp. 223–250. DOI: 10.1080/13600834.2018.1458455. (In Eng.).

26. Omrani N., Rivieccio G., Fiore U., Schiavone F., Garcia-Agreda S. To Trust or Not to Trust? An Assessment of Trust in AI-Based Systems: Concerns, Ethics and Contexts. *Technological Forecasting and Social Change*. 2022. Vol. 181. Article 121763. DOI: 10.1016/j.techfore.2022.121763. (In Eng.).

27. Papavasiliou N. *The Vicissitudes of Law in the Digital Age: Automation as a Mechanism for Justice in Family Law*. Thesis for Bachelor of Laws (Honours). Brisbane, 2020. Pp. 1–5. (In Eng.).

28. John A. M., Aiswarya M. U., Panachakel J. T. *Ethical Challenges of Using Artificial Intelligence in Judiciary*. 2025. Available at: https://arxiv.org/abs/2504.19284. (In Eng.).

29. Salloum S. Trustworthiness of the AI. In: Al-Marzouqi A., Salloum S. A., Al-Saidat M., Aburayya, A., Guta B. (eds.) *Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom*. Springer International Publishing, 2024. Pp. 643–650. DOI: 10.1007/978-3-031-52280-2_41. (In Eng.).

30. Salter S. Online Dispute Resolution and Justice System Integration: British Columbia's Civil Resolution Tribunal. *Windsor Yearbook of Access to Justice*. 2021. Vol. 34 (1). Pp. 112–129. (In Eng.).

31. Sourdin T. Judge v Robot: Artificial Intelligence and Judicial Decision-Making. *UNSW Law Journal*. 2021. Vol. 41. Issue 4. Pp. 1114–1133. (In Eng.).

32. Tolan S., Miron M., Gomez E., Castillo C. Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia. *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*. 2019. Pp. 83–92. (In Eng.).

33. van der Put C. E., Assink M., Boekhout van Solinge N. Predicting Child Maltreatment: A Meta-Analysis of the Predictive Validity of Risk Assessment Instruments. *Child Abuse & Neglect*. 2017. Vol. 73. Pp. 71–88. (In Eng.).

34. Federal Ministry of Justice (Austria). *IT Applications in the Austrian Justice System*, 2025. (In Eng.).

35. Yang Y. Who Should I Trust: Human-AI Trust Model in AI Assisted Decision-Making. *Lecture Notes in Education Psychology and Public Media*. 2024. Vol. 41. Pp. 236–241. DOI: 10.54254/2753-7048/41/20240805. (In Eng.).

36. Zakharova A. V., Staroverova M. S. *Tekhnologii povysheniya professional'noy kompetentnosti pedagogicheskikh kadrov v usloviyakh inklyuzivnoy obrazovatel'noy organizatsii* [Technologies for Raising Professional Competence in Teachers Working in Inclusive Educational Institutions]. *Psikhologicheskaya nauka i obrazovanie PSYEDU.ru* – Psychological Science and Education PSYEDU.ru. 2016. Vol. 8. Issue 3. Pp. 153–165. DOI: 10.17759/psyedu.2016080314. (In Russ.).

37. Završnik A. Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings. *European Journal of Criminology*. 2021. Vol. 18. Issue 5. Pp. 623–642. (In Eng.).

38. Zeleznikow J. Can Artificial Intelligence and Online Dispute Resolution Enhance Efficiency and Effectiveness in Courts. *International Journal for Court Administration*. 2017. Vol. 8. Issue 2. Pp. 30–45. (In Eng.).

39. Zhang Y., Liao V., Bellamy R. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency*. 2020. Pp. 295–305. DOI: 10.1145/3351095.3372852. (In Eng.).

About the author:

S. S. Gulyamov
Tashkent State University of Law
35, Sayilgokh st., Tashkent, 100047, Uzbekistan
ORCID: 0000-0002-2299-2122

Информация об авторе:

С. С. Гулямов
Доктор юридических наук, профессор,
заведующий кафедрой киберправа
Ташкентский государственный юридический университет
100047, Узбекистан, г. Ташкент, ул. Сайилгох, 35
ORCID: 0000-0002-2299-2122