Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

Articles

Studia Iuridica Lublinensia vol. 32, 4, 2023
DOI: 10.17951/sil.2023.32.4.157-173

Yachiko Yamada
Chuo University, Japan
ORCID: 0000-0001-8699-4947
yachiko@tamacc.chuo-u.ac.jp

Judicial Decision-Making and Explainable AI (XAI) – Insights from the Japanese Judicial System

Sądowe podejmowanie decyzji i interpretowalna sztuczna inteligencja (XAI) – spostrzeżenia z japońskiego systemu sądowego

ABSTRACT

The recent development of artificial intelligence (AI) in information technology (IT) is remarkable. These developments have led to claims that AI can be used in courts to replace judges. In the article, the author addresses a matrix of these issues using the concept of explainable AI (XAI). The article examines how regulation can ensure that AI is ethical, and how this ethicality is closely related to (XAI). It concludes that, in the current context, the contribution of AI to the decision-making process is limited by the lack of sufficient explainability and interpretability of AI, although these aspects are adequately addressed and discussed. In addition, it is crucial to consider the impact of AI's contribution on the legal authority that forms the foundation of the justice system, and a possible approach is suggested to consider conducting an experimental study as AI arbitration.

Keywords: explainable AI; XAI; artificial intelligence; courts; judges; decision-making process; judicial decision-making

INTRODUCTION

As a result of the advancement of information technology (IT) in court proceedings, the court uses the latest information technologies. A wide range of software is also in use in the courts, e.g. for case law research. In addition, the recent development of artificial intelligence (AI) in IT is noteworthy. These developments have

CORRESPONDENCE ADDRESS: Yachiko Yamada, PhD, Full Professor, Chuo Law School, 3-11-15 Kandasurugadai, Chiyoda-ku, Tokyo 101-8324, Japan.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

led to claims that AI can be used in the courts to replace judges. The purpose of this article is to examine whether AI will somehow contribute to the judge's legal reasoning or, in some cases, replace a judge.[1]

There is somehow a movement in academia to use AI in courts because of this AI development. Some academics are seriously considering such a move as an AI judge or advocate.[2] Given the capabilities of AI, this may not be so far-fetched after all. However, the products of using AI in the courts are very different from those of using software in general. In the case of general software, such as case law research, the software developer would design how to process the input. This would then be implemented (programmed) by the developer. Artificial intelligence can be self-learning after training data has been fed into the AI.[3] This means that it's impossible for the developer to fully predict what the AI will end up doing. Instead, by learning from the data as it is used, the AI would generally change its own output function. Because of these characteristics, AI algorithms are said to be able to perform much more complex tasks than conventional software.

There are a number of issues that need to be addressed before AI judges can become a reality decision-making process. To what extent can the courts tolerate the contribution of AI? If AI can contribute in some way, will we consider how it can contribute? Or should the state authorise an AI court system if the parties to a dispute are in favour of an AI court? In this article, I would like to address a matrix of these issues using the concept of explainable AI (XAI).[4] This concept has attracted a great deal of attention in recent times. My thesis in this article is grounded in insights derived from the Japanese judicial system. I firmly believe that these insights generally hold relevance to varying degrees in understanding the decision-making process employed by judges beyond Japan case.

REGULATING AI: AN ETHICAL FRAMEWORK

As noted in the introduction, in deep learning AI, a certain amount of data is input and the AI itself trains itself to mimic the content of the data. In such cases, the AI can learn flexible processing in accordance with the input data. In this way, AI can replace tasks related to human intentions and decisions that have been difficult to achieve

[1] It is important to emphasize that there are different types of AI. In this article, I focused not on the so-called expert system, but on a form called machine learning, especially deep learning.

[2] T. Nishimura, *The Possibility of a Vending Machine for Judgment* (*Hanketsu Jidouhanbaiki no Kanousei*), [in:] *Artificial Intelligence Law and Society* (*AI de Kawaru Hou to Syakai*), ed. M. Usami, Tokyo 2020, pp. 137–154.

[3] D. Rothman, *Hands-On Explainable AI (XAI) with Python: Interpret, Visualize, Explain, and Integrate Reliable AI for Fair, Secure, and Trustworthy AI Apps*, Birmingham 2023, pp. 1–2.

[4] *Ibidem*, pp. 3–4.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

with traditional software. There are a great many important tasks that involve the intentions and decisions of a person who has a strong influence on other people. It is also the judiciary that has such a decisive influence on the others. In other words, what is required in a system like adjudication, which has a coercive and decisive influence on others, is that the output results are fair and ethical. Therefore, in order to use AI algorithms in the process of the courts, we must first of all examine whether or not a fair result will be achieved. This can be expressed in terms of 'trustworthy ethics'. The second question that needs to be examined is whether the conclusion presented by the AI is persuasive or convincing. The nature of persuasiveness would focus not on comprehension but on confidence in the AI's algorithms.

1. Ethics for trustworthy AI

How can regulation ensure that AI is ethical? These guidelines are so-called 'soft laws', i.e. laws that are the product of organisations other than the state and do not have the force of law, although they may be the product of the state. To address the ethical issues associated with AI, a number of principles and guidelines for AI have been published. There is considerable overlap in the content of the guidelines. Using the *Ethics Guidelines for Trustworthy AI* by the Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission,[5] I will now review and explain the key principles relevant to the purpose of this paper. In the document, part of the ethical issues includes fairness, transparency and accountability, as well as XAI and interpretable AI. Fairness, accountability and transparency are key principles I will focus on. These are also closely related to the question of explainability.

Fairness.[6] It involves eliminating biases that cause unfairness so that AI can provide fair services regardless of user characteristics. Fairness can be improved by validating the data and AI algorithms from a variety of perspectives. One example that immediately comes to mind of where the fairness of the resulting outcomes is challenged is the issue of bias by racial or gender prejudice. Artificial intelligence based on deep learning learns the relationship between input and output from sample data as it learns, and autonomously acquires processing that mimics it. In other words, if there is a bias in the input data during training, processing will be acquired in accordance with the bias. Such biases include historical bias (historical bias based on people's social conventions in the past) and sampling bias (the use of biased data sources when collecting data). There are various examples of bias, such as users with different attributes being given significantly different scores, or

[5] Independent High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 8.4.2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (access: 10.11.2023).

[6] *Ibidem*, p. 18, para. 1.5 Diversity, non-discrimination and fairness.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

users with certain attributes not receiving the same service as others. This problem of bias is easy to identify. For example, if women are unnaturally excluded from the initial selection to work in AI, it's clear that gender bias exists. Then it is relatively easy to clarify fairness by quickly reviewing data.

Accountability.[7] Accountability means being clear about what went wrong and who is responsible. Artificial intelligence relies on past data and may give incorrect answers depending on past data. However, it is not easy to identify the source of problems in the data used for learning, the algorithms and the overall system design. This is because AI acquires knowledge by learning. In other words, if there is a bias in the input data during training, the processing will be in line with the bias.

In addition, even if a problem is deliberately caused, it will be difficult to hold the company or organisation accountable if there is no clear evidence that this was the case. For AI to be accountable, it needs to show what was seen in the input data that led to the output, the reasons why, and clarify which of the elements that make up the AI system was the cause.

Transparency.[8] Transparency is the ability to present information from AI systems in a way that users can understand. Transparency is the flip side of AI accountability and is necessary for users to have confidence in the use of AI systems. This includes information about what type of data was used for learning, what type of review was conducted, and what criteria and reasons were used for processing. This is important when making decisions, such as in medical assessments, where the impact of a problem is large. Transparency is the flip side of AI accountability and is necessary for users to feel comfortable using the AI system.

Then there is the question of explainability itself as XAI. Explainability is a commonly desired feature among fairness, accountability and transparency. Explainability is distinct from, but closely related to, the three concepts set out in the *Guidelines*. AI systems should be able to explain what processing the AI has learned and on what basis the output was determined for each input.

2. Balancing comprehension and confidence: exploring the relationship between understanding and trust in AI algorithms

The second issue is related to understanding and trusting AI algorithms. It can also be represented as 'comprehension' and 'confidence'. The feature that processing can be acquired automatically through learning means that it is not possible to show exactly how the conclusion was derived. This is going to lead to things that are unacceptable and incomprehensible. This goes hand in hand with the question of explainability of AI. In addition, explainability is closely related to the question

[7] *Ibidem*, p. 19, para. 1.7 Accountability.
[8] *Ibidem*, p. 18, para. 1.4 Transparency.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

of whether AI will be able to fulfil the accountability equivalent to human responsibility when AI replaces human work. Indeed, it has been reported that AI seems to make egregious mistakes.[9] And if it is a mistake that everyone notices, it can be corrected immediately, but if it is a mistake that seems correct at first glance, it will not be discovered immediately and the wrong result will be realised.

Let's not deny that AI can sometimes produce results that are more accurate and more efficient than those produced by humans. However, although they are generally more accurate and faster than humans, they also make mistakes that are impossible for humans to make, and it is not clear why their accuracy is so high. The fact that we do not know why the accuracy is so high is often described as a black box.[10] The answer may vary from person to person as to whether or not to trust the answer that comes out of such a black box. In other words, whether to believe the output of a black box whose judging process is invisible can be said to be a psychological problem on the human side. Should we not dismiss it as a problem of social psychology? For example, how much probabilistic inference that is structurally error-prone in the cost-benefit relationship can be tolerated? How much can the accuracy be increased so that society has the feeling that there is nothing more to be done? There would be a split opinion. Such an opinion, that it is only a psychological problem, might be possible as an ultimate option.

I think there is something we should consider before dismissing it as a psychological problem. In low-influenced areas, overconfidence may not matter. If you think about integrating AI into areas where it has a strong influence, especially in the judiciary, it would be difficult to make such a distinction. It is crucial to consider one key aspect in this context. People will make a distinction between simply being informed about an AI algorithm, i.e. having some understanding of how it works, and trusting the outcomes that are enforced by that AI algorithm. It must be accepted as reasonable to understand as 'comprehension', but not to trust as 'confidence'. It is in this context that the question of the judicial process, which is the subject of our discussion, must be particularly taken into account. The field of XAI is trying to answer this question.

[9] T. Watanabe, *Technological Innovation and Humans – Acceptance of AI* (*Gijyutsu Kakushin to Ningen – AI no Juyou*), [in:] *Artificial Intelligence Law and Society* (*Jinko Chino to Ningen Syakai*), eds. S. Inaba et al., Tokyo 2020, pp. 64–68.

[10] D. Rothman, *op. cit.*, pp. 4–5.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

THE CHARACTERISTICS OF AI KNOWLEDGE AND XAI

Deep learning exhibits a remarkable level of performance; however, the precise mathematical reasons behind its exceptional efficacy remain unclear. Presently, there is extensive discourse regarding the various approaches to elucidate its underlying mechanisms.

1. Algorithmic foundations of deep learning

As mentioned earlier, deep learning exhibits an impressive level of performance; on the other hand, it is frequently characterized as a 'black box'. What does this imply in practical terms? It signifies the challenge of elucidating the inner workings of complex algorithms, like deep learning, in a manner comprehensible to humans. Specifically, it involves articulating the fundamental mechanisms underpinning the algorithm's outputs. Therefore, let us delve into the technical details further to provide a more comprehensive breakdown of this concept.

A multilayer machine learning algorithm, known as the multilayer perceptron (MLP), falls under the category of deep learning neural networks. A perceptron serves as a mathematical abstraction of the neurons found in higher organisms' brains, and its structure bears resemblance to the widely known logistic regression model. Multilayer perceptron is composed of interconnected perceptron arranged in layers. Through the transmission of information across these layers, MLP is capable of capturing complex input-output relationships beyond the capabilities of logistic regression. Nevertheless, incorporating additional parameters into the model results in an increase in the number of learnable parameters within MLP. Furthermore, it is widely recognized that the interpretation and significance of each parameter become increasingly intricate as the layer approaches the output in MLP. The learning outcomes of the layer nearest to the input are more easily comprehensible since they involve direct weighting of specific inputs. However, the learning outcomes of the layer closer to the output entail intricate combinations of weighted inputs, rendering them challenging to interpret.

Multi-layer neural networks, such as MLP, are recognized for their inherent challenge in interpretation due to the significant number of internal parameters. Ongoing endeavors are directed towards enhancing the explainability of these networks. Some people point out that deep learning is extremely special among models that mimic neural networks. As previously mentioned, neural networks constitute the fundamental technology underlying deep learning. Theoretical models, specifically Hidden Markov models in the field of stochastic statistics, can be integrated into neural networks. It appears that an equivalence exists between Hidden Markov models and neural networks, offering valuable insights into the capabilities and limitations of the latter. However, the utilization of deep learning

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

has been accompanied by a dearth of theoretical models, impeding a comprehensive understanding of its underlying mechanisms for achieving favourable results. Nonetheless, practitioners persist in employing deep learning, even in the absence of a well-established theoretical framework. Notably, a Japanese commentator highlighted that the algorithm is perceived as an enigmatic technology, reminiscent of a magical black box. Nevertheless, its continued application in real-world scenarios perseveres.[11]

In the realm of AI, the explainability of a system is contingent upon the complexity of its constituent algorithms. Put simply, as algorithms grow in complexity, they possess a greater capacity to internally represent and emulate intricate reasoning processes in response to input. However, this increased complexity renders the internal representation increasingly challenging to comprehend and articulate. Conversely, simpler algorithms facilitate explainability in AI systems but may encounter difficulties when confronted with complex problem-solving scenarios. Hence, it is imperative to acknowledge the inherent trade-off between the level of complexity and the extent to which an AI system can be explained.

2. Exploring the utilization of XAI

In the quest to achieve both the intricate expressiveness of AI and a higher level of explainability, a technology known as XAI has emerged.[12] It aims to strike a balance between the inherent complexity of AI systems and the imperative need for interpretability. Through the use of transparent and interpretable models, XAI seeks to shed light on the inner workings of AI algorithms, allowing humans to understand and explain the decision-making processes involved to a significant degree.

Explainable AI is an attempt to shed light on the intricacies of complex AI systems from a variety of different perspectives. An illustrative example of how explainability applies to AI research is seen in contexts such as credit screening, where the need for explainability arises. Explaining the rationale behind the decision-making process becomes essential in scenarios where AI algorithms assess the counterparty's ability to repay instead of human decision-makers. This information is invaluable when communicating with customers who may face financial repercussions as a result of AI-driven decisions, or who may express dissatisfaction with

[11] The technical description in this section is mainly based on C. Simon, *Deep Learning and XAI Techniques for Anomaly Detection: Integrate the Theory and Practice of Deep Anomaly Explainability*, Birmingham 2023, pp. 3–26.

[12] The technical description in this section is mainly based on D. Rothman, *op. cit.*, pp. 6–52; N. Ohtsubo et al., *XAI: What Did You Think of Artificial Intelligence Then?* (*XAI – Sonotoki Jinkouchinou ha Dou Kangaetanoka*), Tokyo 2021, pp. 28–43.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

test results. In these cases, effective communication of the underlying processes and factors that influenced the AI-based decisions becomes crucial.

In the field of AI research, there are several approaches that seek to provide insights into XAI from different perspectives.

Firstly, a basic classification is based on the scope of the explanation provided. It includes the global explanations, which explain the behaviour of the whole model, and the local explanations, which explain the reasoning behind individual predictions for each input data. The global explanations seek to understand the AI model itself, highlighting the distinctive values within the model as a whole and identifying influential learning data that strongly influence predictions. Whereas the local explanations aim to interpret the prediction results for individual cases, shedding light on the specific features that played a significant role in the decision-making process.

Second, there is a divergence in the methods used for explanation. Simple calculations of important feature values, visualisation techniques such as decision trees to illustrate decision rules, and approaches that present data with significant impact are all variations of explanation methods.

1. Explanation by feature values. The most straightforward method of explanation is to use feature values. We can represent the influence of these features on certain input data by calculating the degree to which each feature contributes to the prediction. In addition, it is important to note that the AI model can be different depending on the type of data that is being used. It is therefore necessary to provide an understanding of how the feature set is used, tailored to the specific characteristics of the input data, when illustrating the AI model. Furthermore, in the case of learning data, the explanation method focuses on identifying the crucial variables that contribute meaningfully to the predictions. This is achieved by highlighting the importance of specific variables within the data record. On the other hand, when dealing with image data, the explanation method involves illustrating the image regions that play a key role in driving the predictions. By visually representing the influential regions, one can gain some insight into the factors that contribute to the model's decision-making process.

2. Explanation using the amount of judgment. In addition to explaining predictions based on feature values exclusively, explanations in the form of judgment rules can be provided to understand the underlying basis for the prediction. This method is analogous to a decision tree with conditional branching within the model. Using a combination of rules that are understandable to humans, rule-based explanations aim to cover the critical aspects of the AI's prediction.

3. Data volume explanation. Data-based explanation involves the use of AI learning data. When a prediction is made on a given input, the reasoning behind the decision is explained by presenting the learning data that had the most significant impact on the prediction. This approach aims to provide insight into the specific

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

learning data that positively or negatively influenced the AI predictions. By utilising explanations based on learning data, efforts can be directed towards improving the AI system by eliminating learning data that adversely affects predictions.

There is also a distinction based on model dependency. Some AI systems limit the explanation to specific AI models, focusing on deep learning models. These specialised AI systems provide a deep understanding of the structure and algorithms of the underlying AI model. Conversely, there are XAI systems that do not restrict explanations to a specific model type. Explainable AI encompasses different AI models together, allowing for more comprehensive explanations. Importantly, however, XAI may not fully exploit the potential for rational explanations based on the unique structures of individual AI models.

3. Understanding interpretable AI

As a distinction from XAI, I would like to introduce the concept of interpretable AI.[13] The differences between XAI and interpretable AI are as follows: XAI focuses primarily on providing explanations for AI predictions without necessarily requiring a detailed analysis of the internal structure of the AI model. This includes methods that provide extrapolation explanations for black-box AI models, which fall into the realm of XAI. On the other hand, interpretable AI refers to AI systems that have the ability to analyse their internal structure and understand the computational processes that lead to predictions. These AI models can assess how predictions are affected by changes in parameters or variations in input data. Classical machine learning methods, such as decision trees, are examples of interpretable AI because they allow the computational process leading to predictions to be traced. Although not directly covered in this paper, it is important to recognize that interpretable AI includes machine learning techniques such as decision trees that facilitate a transparent understanding of the computational processes leading to predictions.

4. Characteristics of AI knowledge: exploring the nature of AI insights

Of course, at least for now, the intelligence generated by AI and deep learning does not exactly match human intelligence. The resulting processes rely heavily on rules of thumb. However, there is a tendency to downplay counter-evidence, resulting in complex and incomprehensible knowledge. It is clear that AI knowledge has a distinct character, which can lead to significant fears and concerns among AI users. For example, there is a risk of over-reliance on biased heuristics due to data bias. There is also the potential to lapse into pseudoscience, making claims about universal laws that cannot be reliably derived from the data alone. The task of

[13] N. Ohtsubo et al., *op. cit.*, pp. 28–29.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

unravelling such complex and enigmatic knowledge is daunting. If the knowledge of AI is heterogeneous, it can be expected to have a significant impact on legal reasoning, which is the central theme of this article.

UTILIZATION OF AI ALGORITHMS IN JUDICIAL DECISION-MAKING

Traditionally, judicial decision-making involves judges adjudicating disputes and settling disputes. The decision of this judge consists of a legal reasoning. It is made in the form of a so-called legal syllogism. Legal syllogisms consist of the application or interpretation, including application, of law, and fact-finding. Moreover, the relationship between the application of law and the determination of facts is so complicated that the application of law, the interpretation and the determination of facts are interrelated in their own way and cannot be separated in the human mind. However, when AI takes over the work of judges, at least for the time being, we should separate fact-finding from interpretation of the application of the law.

In this chapter, I will explore the potential of AI to enhance legal reasoning within the judicial decision-making process while considering the fundamental principles of the rule of law. Explainability of AI judgments is a crucial and discussed topic,[14] as previously mentioned. At the same time, it is worth noting that certain individuals argue that the decision-making process of human judges in judicial contexts is also considered a black box. Some theorists argue that because the decision-making process of human judges is also considered a black box, there should be no issue with the decision-making process derived from AI being a black box as well. However, it is crucial to critically examine this perspective. Even though both the judgment of a human judge and the judgment of an AI are often referred to as black boxes, it is crucial to acknowledge that the nature of knowledge generated by an AI differs from that of human thought. Therefore, it is essential to rigorously distinguish and address these differences. There are two critical issues that require careful consideration regarding the use of AI in judicial decision-making. Firstly, the question arises as to whether a sufficient volume of data can be obtained. Secondly, there is the concern of whether the tasks performed by professionals can be effectively replaced. In this article, I will examine each of these issues in detail.

[14] A. Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, "Columbia Law Review" 2021, vol. 119(3), pp. 1829–1830.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

1. Data sufficiency: a legal perspective

Let's begin by addressing the first issue. As previously mentioned, deep learning algorithms require a substantial amount of data to make accurate decisions. Therefore, it is crucial to ensure an adequate volume of data is available to support the AI system's decision-making process. Securing large amounts of data that can be utilized by AI algorithms is not always a straightforward task. However, the emergence of Wikipedia in 2001 has significantly transformed the landscape. The outcome is an extensive information resource consisting of over a million entries in 14 languages. Although Wikipedia itself does not employ the technical components of AI, it serves as an invaluable resource for AI by providing the requisite volume of data. Since then, the large amount of data on the Internet has played an important role in the utilization of AI, including generative AI.

What about judicial decision-making? There is a growing demand for access to judgments and court records. The digitization of judgments and court records varies across countries. In Japan, the digitization of judgments and court records has made limited progress. In 2017, the Cabinet Secretariat established a study group to explore the implementation of IT in judicial proceedings.[15] However, the level of computerization has not yet reached an adequate stage, and currently, only a small number of judicial precedents have been digitized.[16] Among these, a limited number of judgments have been made accessible to the public. Furthermore, the majority of published judgments pertain to cases involving disputed interpretations of legal texts, while judgments on general cases that establish legal conclusions have not been released. A Japanese commentator also has highlighted that the publication of all current judgments alone does not guarantee accurate decision-making through deep learning, as the available amount of data is insufficient.[17]

2. Expertise in the legal judgment system: a comprehensive analysis of judicial decision-making and legal education

The next aspect I would like to address is the expertise within the legal judgment system. Judges are acknowledged as experts, which leads us to consider the possibility of AI replacing or augmenting their expertise and workload.

---

[15] Research Group on the Introduction of Information Technology in Judicial Procedures, Promoting the Use of IT in Judicial Proceedings (Future Investment Strategy 2018), Cabinet Decision of 15 June 2018), https://www.kantei.go.jp/jp/singi/keizaisaisei/saiban/index.html (access: 10.11.2023).

[16] Courts in Japan, https://www.courts.go.jp/app/hanrei_jp/search1 (access: 10.11.2023).

[17] I. Sato, *The Communication between Technology and the Laws* (*Tekunologii to Hou no Taiwa*), [in:] *Transformation of the Laws under Society with Artificial Intelligence* (*AI to Syakai to Hou – Paratdaimushihuto ha Okiruka*), eds. J. Shishido et al., Tokyo 2020, pp. 5–6.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

In any legal system, judges typically possess specific qualifications that enable them to serve as judges. Similarly, legal professionals such as lawyers, who represent parties in litigation, are also required to meet certain qualifications. The requirements for obtaining these qualifications typically involve a combination of educational degrees, examinations, and professional training, although the specific criteria may vary across different jurisdictions and over time. Regarding the legal profession, in certain countries, it is possible to qualify for legal practice without the requirement of an examination.

Some countries, like the United Kingdom, have established legal education systems with a long history,[18] whereas Japan in Asia adopted its legal education system relatively recently, during the Meiji era in the 19th century.[19] Moreover, in the case of legal professionals, the authority responsible for granting qualifications can be either governmental or non-governmental (such as legal professional organisations). Judges, on the other hand, are appointed and qualified by the state. The qualifications for judges differ conceptually from those for attorneys, and there are countries where the legal profession is unified, as well as countries where it is not. In the UK, for instance, it is generally required to have experience as a barrister in order to become a judge, granting barristers the opportunity to pursue judicial positions. In Japan, individuals who aspire to become judges, prosecutors, and attorneys. After successfully passing an examination to enter a training center for legal professionals, they proceed to the Legal Training and Research Institute, which operates under the jurisdiction of the Supreme Court. In modern legal systems, it is crucial to have the ability to qualify as a judge through a combination of a degree, an exam, and subsequent training.

In the case of Japan, aspiring judges typically begin their careers as assistant judges and gradually progress to become judges. Typically, after approximately 10 years of service as an assistant judge, they have the opportunity to become independent judges. Judges undergo a rigorous training process to develop their expertise and acquire the necessary skills for making judgments.

If AI is capable of making judgments on behalf of expert judges, it raises questions about the algorithmic requirements needed to replace the expertise acquired through extensive training. Is such a replacement necessary? If AI can supplant the training curriculum required to become a specialist, it may also cast doubt on the identity and existence of the legal profession itself. When examining this issue, it is crucial to assess the extent to which AI judgments contribute to the process of judicial decision-making.

If AI is to completely replace expert judges, it is a very serious problem. But even if AI does not completely replace expert judges, but merely contributes in

[18] P. Darbyshire, *English Legal System*, London 2020, pp. 301–302.

[19] H. Oda, *Japanese Law*, Oxford 2009, pp. 73–74, 84–85.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

some way to the legal reasoning leading to judicial decisions, it still has to be taken seriously. This won't work. If AI were to completely replace expert judges, it would pose a significant and concerning issue.

However, even if AI only contributes to the legal reasoning process leading to judicial decisions, it must be treated with utmost seriousness. The reason for this lies in the potential magnitude of harm caused by an undesirable outcome resulting from an inappropriate decision made by a judge. This magnitude far exceeds the consequences of a member of the public or a student receiving an inappropriate answer from a generative AI when posing a legal question. It is crucial to consider the fact that AI, unlike a human judge, lacks the ability to explicitly articulate the reasons behind its judgments. This raises the question of whether AI algorithms can ensure compliance with the requirement for judicial decisions to allow for review through the three-instance system.

The written judgment not only includes the final outcome (win or lose) but also provides the rationale behind the decision. The legality of the decision is examined on the basis of the reasons given in the judgment. A party may bring an action against the judgment in the form of an appeal. In Japan, it is not uncommon for a first instance judgment to be overturned on appeal. Review of the correctness of the judge's decision may take the form of judicial commentary, more commonly by academics. So can AI algorithms be used to ensure such verification mechanisms? It is difficult to imagine a first-instance AI decision being overturned by a second-instance AI.

3. Distinguishing the characteristics of knowledge in judicial decision-making: on black box of judicial decision-making

As I mentioned in the previous section, when examining judicial decision-making, the issue arises of whether human judges, as experts in their field, can be replaced. When discussing judicial decision-making, it is important to consider the expertise of government officials and the three-court system. Additionally, scholars have highlighted the opaque nature of the legal thinking employed by human judges. To describe this opaque nature, the term 'black box' is sometimes used.

As previously mentioned, AI decision-making is often characterized as operating within a black box, given the inherent opacity resulting from the processing of vast amounts of complex data using sophisticated algorithms. I elucidated the introduction of the concept of XAI as a potential solution to address the challenges associated with the inherent black-box nature of AI. It is important to note that XAI is an ongoing area of development and research.

To begin with, it is essential to delve into the inherent characteristics of the black-box nature of AI judgment. To a certain degree, it is widely acknowledged that the output generated by a machine, including calculators and computers, can

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

be perceived as a black box, concealing the intricacies of its calculation processes. However, it is crucial to note that even complex pieces of equipment are built upon the premise that there exists a comprehensive understanding of each constituent part. This distinction becomes particularly evident in the context of deep learning. Within deep learning, the intricate nature of neural networks presents a challenge as the specific operations performed at each step in the post-learning information processing phase, as well as their cumulative impact on the overall process, remain uncertain. Furthermore, deep learning models require a vast amount of training data, and it is impractical for humans to manually inspect, classify, label, and evaluate the entirety of this information. That's why it's said that no one understands exactly how AI systems learn to make the decisions they make today.

In considering the argument that AI is often referred to as a black box, it is important to note that the legal reasoning by the human judges themself can also be seen as a black box. While AI systems may be criticized for their opacity, the decision-making processes within the judiciary are often complex and not easily understood. The role of human judges in court rulings is significant, but there is ongoing debate regarding the extent to which the logical reasoning employed by judges in the legal decision-making process is the sole determinant of their judgments. If we consider that judges derive deductive conclusions by interpreting legal texts and identifying relevant facts, it can be argued that the logical reasoning presented in their judgments provides a comprehensive explanation of their decisions. However, an alternative perspective suggests that in addition to the logic explicitly stated in judgments, a judge's intuition and experience may play a role in reaching a conclusion. According to this view, the reasons provided in a judgment serve as a justification for the decision made after reaching the conclusion. If we consider the notion of judicial decision-making by human judges as a black box, it encompasses the aspects previously discussed.

What is the relationship between the opaqueness of the decision-making process in human judges, often referred to as a black box, and the opaqueness of AI decision-making, also characterized as a black box? While the characteristics of these two opaques may appear similar on the surface, their internal mechanisms are fundamentally different. In the realm of deep learning, AI utilizes vast amounts of data to identify patterns, establish correlations between certain characteristics and corresponding categories, and subsequently deduce judgments along with their underlying rationales. However, this information is represented through the configuration of neural networks and the assignment of numerous parameters, which are not readily interpretable by humans in the form of language or easily understandable formulas. In some cases, only the conclusion is provided without any insight into the reasoning behind it. To address this issue and shed light on the decision-making process, a solution called XAI has been proposed, which aims to visualize and explain how an AI system arrives at its decisions. In some cases,

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

only the conclusion is provided without any insight into the reasoning behind it. The distinction between AI and human judges becomes evident when considering the differences in their thinking processes and methods of verification. In the context of generalization beyond the field of law, it has been argued that machines are incapable of achieving the level of mathematical discoveries accomplished by mathematician H. Poincaré.[20]

Based on the preceding discussion, it becomes evident that while the term 'black box' is employed to describe both human judges and AI systems, they exhibit fundamentally distinct characteristics. While it may be tempting to assume that the use of identical terminology implies similar content, a closer examination reveals that such an assumption can be misleading. The mere use of the term 'artificial intelligence' does not imply equivalence to human intelligence. Similarly, when referring to AI judges, it is crucial to avoid misconceptions that envision trials conducted solely by AI-equipped judges. While my discussion thus far has focused on legal reasoning, it is important to note that the scope does not encompass fact-finding, which is beyond the scope of this paper.

4. Authority of judicial system

As mentioned earlier, the legal reasoning employed by human judges is occasionally characterized as a black box. However, it is crucial to recognize that this characterization reflects the intricate and uniquely human thought process, which cannot be entirely captured in the written rationale behind a judgment. However, despite the possibility of the reasons for a judgment not being fully articulated, it is important to acknowledge that the actions of judges, who are institutionally recognized as experts, and the justifications provided in their judgments can be subject to scrutiny, review, and potential overturning on appeal or by a third party. It can be argued that these mechanisms of scrutiny and review of judicial actions and the justifications provided in judgments serve to uphold the authority of the judicial system.

CONCLUSIONS

As described above, whether the introduction of such AI (using deep learning) can contribute to the legal reasoning process of the courts has been questioned should be cautious. In the current context, the contribution of AI in the decision-making process is limited due to the lack of sufficient explainability and interpretability. Until these aspects are adequately addressed and discussed, the

[20] M. Kureha, M. Kukita, *AI and Science Research* (*AI to Kagakukenkyuu*), [in:] *Artificial Intelligence Law and Society* (*Jinko Chino to Ningen Syakai*), eds. S. Inaba et al., Tokyo 2020, p. 142.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

potential of AI to contribute significantly remains limited. Additionally, it is crucial to consider the impact of AI's contribution on the legal authority that forms the foundation of the justice system.

One possible approach is to consider conducting an experimental study. The introduction of AI into the courts may raise concerns regarding its reliability and potential consequences for public trust in the judiciary. Some individuals may argue that a cautious approach is necessary, as a failure or misuse of AI in the judicial process could undermine confidence in the judicial system. Therefore, the concept of AI-driven arbitration is also an intriguing proposition. Arbitration, unlike mediation, is a private dispute resolution mechanism that does not involve state institutions. The idea of employing AI in the arbitration process opens up possibilities for leveraging advanced technologies to facilitate impartial and efficient resolution of disputes. In the context of AI-assisted arbitration, the existence of multiple parties involved in the arbitration process may lead to a diverse range of options available in the market. However, it is important to note that arbitration awards are typically enforceable, highlighting the need to ensure adequate procedural safeguards when incorporating AI technologies in the arbitration process. This ensures that the outcomes produced through AI-assisted arbitration are reliable, fair, and maintain the necessary level of enforceability.

REFERENCES

Literature

Darbyshire P., *English Legal System*, London 2020.

Deeks A., *The Judicial Demand for Explainable Artificial Intelligence*, "Columbia Law Review" 2021, vol. 119(3).

Kureha M., Kukita M., *AI and Science Research* (*AI to Kagakukenkyuu*), [in:] *Artificial Intelligence Law and Society* (*Jinko Chino to Ningen Syakai*), eds. S. Inaba et al., Tokyo 2020.

Nishimura T., *The Possibility of a Vending Machine for Judgment* (*Hanketsu Jidouhanbaiki no Kanousei*), [in:] *Artificial Intelligence Law and Society* (*AI de Kawaru Hou to Syakai*), ed. M. Usami, Tokyo 2020.

Oda H., *Japanese Law*, Oxford 2009, **DOI: https://doi.org/10.1093/acprof:oso/9780199232185.001.1**.

Ohtsubo N. et al., *XAI: What Did You Think of Artificial Intelligence Then?* (*XAI – Sonotoki Jinkouchinou ha Dou Kangaetanoka*), Tokyo 2021.

Rothman D., *Hands-On Explainable AI (XAI) with Python: Interpret, Visualize, Explain, and Integrate Reliable AI for Fair, Secure, and Trustworthy AI Apps*, Birmingham 2023.

Sato I., *The Communication between Technology and the Laws* (*Tekunologii to Hou no Taiwa*), [in:] *Transformation of the Laws under Society with Artificial Intelligence* (*AI to Syakai to Hou – Paratdaimushihuto ha Okiruka*), eds. J. Shishido et al., Tokyo 2020.

Simon C., *Deep Learning and XAI Techniques for Anomaly Detection: Integrate the Theory and Practice of Deep Anomaly Explainability*, Birmingham 2023.

Pobrane z czasopisma Studia Iuridica Lublinensia **http://studiaiuridica.umcs.pl**
Data: 13/04/2025 12:42:02

Ward J., *Black Box Artificial Intelligence and the Rule of Law*, "Law & Contemporary Problems" 2021, vol. 84(3).

Watanabe T., *Technological Innovation and Humans – Acceptance of AI* (*Gijyutsu Kakushin to Ningen – AI no Juyou*), [in:] *Artificial Intelligence Law and Society* (*Jinko Chino to Ningen Syakai*), eds. S. Inaba et al., Tokyo 2020.

Online sources

Courts in Japan, https://www.courts.go.jp/app/hanrei_jp/search1 (access: 10.11.2023). [in Japanese]

Independent High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 8.4.2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (access: 10.11.2023).

Research Group on the Introduction of Information Technology in Judicial Procedures, Promoting the Use of IT in Judicial Proceedings (Future Investment Strategy 2018), Cabinet Decision of 15 June 2018), https://www.kantei.go.jp/jp/singi/keizaisaisei/saiban/index.html (access: 10.11.2023). [in Japanese]

ABSTRAKT

Niedawny rozwój sztucznej inteligencji w technologii informacyjnej jest niezwykły. Zmiany te doprowadziły do twierdzeń, że sztuczna inteligencja może być wykorzystywana w sądach do zastępowania sędziów. W artykule autor odnosi się do sedna tych problemów, używając koncepcji interpretowalnej sztucznej inteligencji (XAI – *Explainable Artificial Intelligence*). Analizie poddano to, w jaki sposób regulacja może zapewnić, że sztuczna inteligencja będzie etyczna, a także w jaki sposób ta etyczność jest ściśle powiązana z XAI. Stwierdzono, że obecnie wkład sztucznej inteligencji w proces decyzyjny jest ograniczony przez brak wystarczającej możliwości jej wyjaśnienia i interpretacji, chociaż aspekty te są odpowiednio uwzględnione i omówione. Ponadto kluczowe jest rozważenie wpływu sztucznej inteligencji na autorytet prawny, który stanowi podstawę wymiaru sprawiedliwości. Zasugerowano przy tym rozważenie przeprowadzenia badania eksperymentalnego polegającego na włączeniu sztucznej inteligencji do procesu arbitrażowego.

Słowa kluczowe: interpretowalna sztuczna inteligencja; XAI; sztuczna inteligencja; sądy; sędziowie; proces decyzyjny; podejmowanie decyzji sądowych