THESES OF DOCTORAL (PhD) DISSERTATION

The Illusion of Consistency:
Examining a Legal Application of Generative AI

dr. Kardos Vivien Kata

Supervisor:
Dr. habil. Kovács Péter, PhD

Szeged, 2025.

Table of Contents

Table of Contents 1

I. Introduction 3

I.1. Reasons for choosing this topic 3

I.2. Purposes and research questions of the PhD dissertation 4

I.3. Research methodology 4

I.4. Hypotheses of the PhD dissertation 5

I.5. Structure of the PhD dissertation 5

II. Theoretical overview 7

II.1.1. The concept of AI 8

II.2. AI and legal work 8

II.3. The development of AI – from its beginnings to the present day 9

III. Some aspects of AI regulatory issues 11

IV. AI literacy and related literacies 15

IV.1. AI literacy in th AI Act 15

IV.2. The scientific approach to AI literacy 15

V. Current trends in the use of AI in the legal work 18

VI. AI-supported applications in legal work 21

VII. Legal LLM benchmarks 23

VII.1. General evaluation metrics 23

VII.2. Main findings of the legal LLM benchmarks 23

VIII. Empirical research 25

VIII.1. Methodological framework 25

VIII.2. The uniqueness of the exploratory research 28

VIII.3. Preliminary examination 29

VIII.3.1. Methodology of the preliminary examination 29

VIII.3.2. Some results of the preliminary examination 31

VIII.4. Empirical research 32

VIII.4.1. Reasons for conducting empirical research 32

VIII.4.2. Methodology 32

VIII.4.3. The questionnaire used as the basis for the empirical research 33

VIII.4.4. Analytical aspects of empirical research 33

VIII.4.5. Generative AI systems used in empirical research 34

VIII.4.6. The applicability of the examined generative AI systems in legal work 35
VIII.5. Findings 35
IX. Research and application limitations 37
X. Results, recommendations 38
XI. Publications on the topic of the PhD dissertation 41

I. Introduction

In recent years, the importance of data has increased, and the approach that data is the new oil has emerged.[1] As the years have passed, this approach has been further confirmed by the increasingly widespread use of artificial intelligence (hereinafter referred to as: AI). The rise of generative AI in everyday use has played a significant role in this. Although AI and generative AI are not new in certain professional and scientific circles, the release of OpenAI's ChatGPT 3.5 model in November 2022[2] and its subsequent explosive growth have taken generative AI and its applications to a new level. In the case of ChatGPT, the number of users exceeded one million in just five days, with 400 million weekly users in February 2025 and 700-800 million in October 2025.[3] The explosion of this technological solution has brought AI and generative AI into general public, even though they had already been present in many areas of everyday life, such as mobile phone camera functions, the 'smart home' concept, and the algorithms behind personalised advertising.

It is important to note that AI and generative AI are having an increasing impact on legal science and legal practitioners, both from a theoretical and practical perspective. In this PhD dissertation, legal work is primarily limited to the work and activities performed by attorneys-at-law and in-house counsels (corporate lawyers).

I.1. Reasons for choosing this topic

AI is suitable for performing a supporting function in the field of legal work, and numerous AI-based solutions are being applied in practice. Based on surveys and trends, this is expected to increase both internationally and domestically. The use of general-purpose generative AI systems is significant in legal work. After a long preparatory process, the European Union has adopted comprehensive regulations and a legal framework for AI, which, due to its regulatory nature, is also mandatory and directly applicable in Hungary. When using AI, it is of paramount importance that individuals have the knowledge and skills that enable the responsible use of AI. AI literacy is also playing an increasingly important role in the labour market.

It is important to note that text analysis, text comprehension and accuracy play a key role in legal work. The specific structure, language and terminology used in legal provisions pose a challenge for generative AI systems, as the omission of even a comma or a word (e.g. 'at least') can give the generated text a completely different meaning. As a result, legal provisions provide a good basis for identifying common errors in generative AI systems. The labour law institution of paternity leave, which is detailed in the exploratory research, and the two legislative amendments affecting it, which came into force at different times, proved to be suitable for examining the text generation solutions of certain generative AI system models in relation to specific legal issues from a practical perspective. My personal reasons

[1] Leaders: The world's most valuable resource is no longer oil, but data, The Economist, 6 May 2017, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [19 October 2024]

[2] OpenAI: Introducing ChatGPT, 2022. november 30., https://openai.com/index/chatgpt [2025. 04. 17.]

[3] Duarte, Fabio: Number of ChatGPT Users (November 2025), last updated: 31 October 2025, https://explodingtopics.com/blog/chatgpt-users [19 November 2025]

for choosing this topic include the fact that, on the one hand, I have dealt with legal IT trends and their significance in my previous research, and on the other hand, I have been able to see for myself the potential of general and legal-specific generative AI systems in a practical approach.

I.2. Purposes and research questions of the PhD dissertation

AI and generative AI are developing at an extremely rapid pace, but it is also clear that these systems can make mistakes, possibly creating the impression that non-existent information exists and corresponds to reality, i.e. they hallucinate.

The purposes of this PhD dissertation include presenting current AI trends, areas of application and applied solutions relevant to legal work. Another aim is to describe the increasingly important AI littracy and other related literacies. Based on the results of specific exploratory research, I intend to present, through a case study, a practical approach to the experiences arising in connection with various generative AI systems in a legal context. I will place particular focus on how the generative AI systems under investigation respond to the legal questions I have identified, what types of errors occur frequently, and whether the developments of the models under examination have resulted in progress in relation to these issues over time. Given that this field is changing extremely dynamically, I aim to provide a snapshot of the current state based on the results.

The research questions of this PhD dissertation were defined as follows:

a) How effectively are the generative AI systems examined able to track legislative changes? Is there a best-performing system among them?
b) How consistent are the responses generated by the generative AI systems under examination, and is there any contradictory content between the responses?
c) How identifiable and consistent is the source of information used by generative AI systems? Are the examined generative AI systems equipped with web search tools capable of better answering simple, specific questions related to legal provisions based on information collected from freely accessible legal database interfaces through the application of web search tools?
d) What types of errors occur in the generative AI systems examined in relation to specific legal provisions?
e) How and to what extent did the generative AI systems examined develop between the preliminary examination and the empirical research?

I.3. Research methodology

From a research methodology perspective, this PhD dissertation consists of two main parts: a literature review – including the issues of AI regulation – followed by a detailed presentation of my exploratory research through case studies.

In reviewing the literature, I used both domestic and international sources, with a greater emphasis on international literature and other professional websites due to the nature of the topic.

The exploratory research and the resulting case study are based on my preliminary examination and empirical research, which involved comparative text analysis. During the preliminary examination and empirical research, I used questionnaires compiled according to criteria I had defined. The individual subchapters describe in detail which questions and which generative AI systems were used during the preliminary examination and empirical research.

I.4. Hypotheses of the PhD dissertation

In connection with the research underlying this PhD dissertation, the following assumptions and hypotheses were formulated.

(1) The generative AI systems examined can be used to track legislative changes, but only under supervision; selecting the appropriate time frame to answer the question is difficult, and in general it is not possible to clearly identify the best system.
(2) Despite developments, the responses of the AI systems examined are still inconsistent, and contradictory content is generated even within a given response.
(3) The generative AI systems examined that have a web search tool are able to answer simple, specific questions related to legal provisions based on information collected from a free legal search database using the web search tool. The generative AI systems examined use the appropriate source from among the sources displayed by the web search engine to provide answers, thereby giving better answers.
(4) For the examined AI systems, frequently occurring error types are identified as: responses based on an incorrect temporal state of legal provisions; responses based on a previous, no longer valid temporal state; and incompleteness in responses to questions requiring a list.
(5) As a result of the developments affecting the generative AI systems examined, the quality and composition of the responses improved overall during the period between the preliminary examination and the empirical research, there are fewer errors in terminology, measurement and dates, and the responses are more accurate and to the point in all of the AI systems examined compared to the preliminary examination.

I.5. Structure of the PhD dissertation

In addition to introductory remarks, the first chapter of this dissertation presents the rationale for choosing the topic, the objectives of the dissertation, the research questions, the research methodology and the hypotheses.

The second chapter presents the development of AI from its beginnings to the present day, regulatory issues surrounding AI, and some key concepts related to AI. The third chapter outlines the milestones in AI regulation in the European Union. The fourth chapter deals with AI literacy and related skills. The fifth chapter presents current trends in the use of AI in legal work. The sixth chapter deals with AI-supported applications used in legal work and areas of AI application. The seventh chapter presents benchmarks related to large language models in the legal field.

The eighth chapter contains the results of empirical research: the main results of preliminary examination and empirical research, findings and conclusions are presented. Chapter 9 identifies the limitations of research and application. Chapter 10 contains the summary and recommendations. This is followed by the bibliography, appendices and annexes, which include the preliminary examination and the questions underlying the empirical research.

This PhD dissertation does not aim to present all the risks and dangers associated with AI.

II. Theoretical overview

The theoretical part of the dissertation consists of seven chapters. As part of the literature review, I presented the conceptual foundations of AI and defined the meaning of certain terms used by me (e.g. legal work, AI use, tracking legislative changes). I then presented the development of AI from the 1940s to the present day, with a particular focus on generative artificial intelligence systems (hereinafter: generative AI systems) that are significant from the perspective of empirical research. Then I described the regulatory framework for AI, including an international overview of the regulatory approaches of the European Union, the United States of America and China. This chapter included a description of the legislative process leading to the adoption of Regulation (EU) 2024/1689 of the European Parliament and of the Council[4] , hereinafter reffered to as: AI Act), which is binding and directly applicable in Hungary, as well as highlighting current developments relevant to the PhD dissertation.

It is important to note that Article 3(1) of the AI Act explicitly defines the concept of an AI system, while Article 3(56) defines the concept of AI literacy. In the chapter on AI Act, I also discuss certain milestones in national AI regulation, including certain topics of the renewed national AI strategy[5]. In connection with the use of generative AI systems, particular importance can be attached to AI literacy, and so its scientific approach and certain obligations set out in the AI Act are also presented. Digital literacy, privacy literacy, algorithmic literacy and data literacy, which are related to AI literacy, are also discussed.

AI is having an increasing impact on legal work, so the relevant literature, international trends, areas of application of AI in legal work and specific AI-supported applications, as well as the examination of benchmarks used in the case of large language models (LLMs) played a prominent role in the PhD dissertation.

As a result, a separate chapter was devoted to reviewing international surveys and reports that examined the use of AI by legal professionals and the impact of AI on the legal profession. In this context, I presented the results of 12 international surveys and reports in detail, including current AI trends. I described AI-supported applications in legal work by area of application (e.g. legal research, document management, document review, AI agents, e-discovery). With regard to legal LLM benchmarks, I first presented the general evaluation metrics, followed by six legal LLM benchmarks (LegalBench, LawBench, LEXTREME, LexGLUE, SCALE, LBOX-OPEN).

For the PhD dissertation, I used a total of 145 references, 217 internet sources, 27 documents that are legal acts, communications, or other documents of the European Union institutions, and 15 Hungarian legal sources. The results of each part of this literature review are presented below.

---

[4] REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Official Journal of the European Union, 12 July 2024.

[5] Magyarország Mesterséges Intelligencia Stratégiája (2025-2030), 2025. szeptember 3., pp. 1-119., p. 52. https://cdn.kormany.hu/uploads/document/c/c0/c0d/c0dfdbd37cfa520ae37361a168d244c85e7295af.pdf [30. October 2025.]

II.1.1. The concept of AI

Considering the development of AI, it is important to note that there are several approaches to defining the concept of AI, but there is currently no uniform, generally accepted definition. The aim of this PhD dissertation is not to examine and present this conceptual area in detail, but to present a few definitions of AI from different perspectives as examples.

From a technological perspective, Stuart Russell and Peter Norvig distinguished four approaches to the concept of AI: (1) thinking humanly, (2) acting humanly, (3) thinking rationally, and (4) acting rationally, in which the dimensions of thinking vs. acting and human vs. rational performance measurement appear.[6] In their study, Zsolt Czékmann et al. present AI as a field of expertise. They define AI as a field of expertise that deals with solving tasks requiring human intelligence through information and communication tools.[7]

According to Article 3(1) of the AI Act: *'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.*[8] Despite the fact that the concept has now been declared at 'European Union level', no general consensus has been reached.

In their study, Ran He et al. defined the concept of generative AI as follows: generative AI refers to a group of AI algorithms and models that are capable of creating new content, including text, images, videos and problem-solving strategies, with human-like creativity and adaptability.[9]

In summary, it can be concluded that one of the key elements of the umbrella term AI is that it refers to a machine system that has a certain degree of autonomy and is used to achieve a specific goal. Furthermore, another important conceptual element is that it replicates human capabilities in some way. In this PhD dissertation, the term AI refers to the concept of an AI system as defined in the AI Act, with the case study presented in PhD dissertation focusing on generative AI systems within this conceptual category.

II.2. AI and legal work

Legal work involves complex professional activities primarily performed by persons with a legal qualification, covering a number of work processes related to the legal profession. In this context, this PhD dissertation focuses on certain activities and tasks

[6] Russell, Stuart – Norvig, Peter: Artificial Intelligence A Modern Approach Third Edition, Pearson, 2010, pp. 1-1132., https://people.engr.tamu.edu/guni/csce625/slides/AI.pdf [2025. 05. 08.] pp. 1-2.
[7] Czékmann Zsolt – Kovács László – Ritó Evelin: Mesterséges intelligencia az államigazgatásban, In: Török, Bernát; Ződi, Zsolt (szerk.) A mesterséges intelligencia szabályozási kihívásai: Tanulmányok a mesterséges intelligencia és a jog határterületeiről, Budapest, Magyarország: Ludovika Egyetemi Kiadó, 2021. pp. 387-402., p. 388.
[8] AI Act, Article 3(1)
[9] Ran, He – Jie, Cao – Tieniu, Tan: Generative artificial intelligence: a historical perspective, National Science Review, Volume 12, Issue 5, May 2025. DOI: https://doi.org/10.1093/nsr/nwaf050

related to the legal profession and performed in legal departments. It is indisputable that AI is having an increasing impact on the field of law, legal work and the legal profession as a whole. It is important to note that, at its current level of development, AI is not yet capable of performing and solving complex legal tasks independently, which is why professional supervision is essential.[10] At the same time, it can provide 'support' in specific tasks, the number and diversity of which are constantly growing.

Several factors contribute to the widespread use of AI in the legal sector, such as the size of the legal market, its characteristics, the regulatory environment and economic development. The impact of AI on the legal profession can be defined as encouraging legal professionals to acquire new skills and, to a certain extent, 'forcing' them to keep up, as developing AI literacy is now essential. It should also be emphasised that the responsibility for the results obtained through the use of an AI system and the use of the content generated by the AI system lies with the lawyer using it.

II.3. The development of AI – from its beginnings to the present day

Although AI has attracted enormous attention in recent years, it is by no means a new technological solution in academic circles, as its roots date back to the 1950s.

Researchers who sought to understand how the brain works played a significant role in the development of AI. In this context, it is worth mentioning that as early as 1943, Warren S. McCulloch and Walter Pitts described in their study that due to the 'all or nothing' nature of neural activity, 'neural events' and the connections between them can be treated with propositional logic.[11] It is important to note that deep learning, which is highly relevant today, is based on the application of neural networks, on the statistical models that these co-authors created based on biological neural networks.[12] The earliest attempts to create an 'artificial brain' led to the invention of the Neumann architecture, named after János Neumann (John von Neumann), who was a Hungarian-born person.[13]

---

[10] Compare: Charlotin, Damien: AI Hallucination Cases, https://www.damiencharlotin.com/hallucinations/?page=2&page=3&page=4&page=3&page=2&page=1&page=2&page=3&page=4 [23 November 2025]
The website contains legal decisions in which the court or tribunal explicitly stated (or implicitly assumed) that one of the parties relied on hallucinatory content or material.

[11] McCulloch, Warren S. – Pitts, Walter: A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 1943. pp. 115–133. p. 115. DOI: https://doi.org/10.1007/BF02478259

[12] Eszteri Dániel: A gépek adatalapú tanításának megfeleltetése a GDPR egyes előírásainak, In: Török, Bernát; Ződi, Zsolt (szerk.) A mesterséges intelligencia szabályozási kihívásai: Tanulmányok a mesterséges intelligencia és a jog határterületeiről, Budapest, Magyarország: Ludovika Egyetemi Kiadó, 2021. pp. 187-210., p. 191.

[13] Zador, Anthony – Escola, Sean – Richards, Blake – Ölveczky, Bence – Bengio, Yoshua – Boahen, Kwabena – Botvinick, Matthew – Chklovskii, Dmitri – Churchland, Anne – Clopath, Claudia – DiCarlo, James – Ganguli, Surya – Hawkins, Jeff – Körding, Konrad – Koulakov, Alexei – LeCun, Yann – Lillicrap, Timothy – Marblestone, Adam – Olshausen, Bruno – Pouget, Alexandre – Savin, Christina – Sejnowski, Terrence – Simoncelli, Eero – Solla, Sara – Sussillo, David – Tolias, Andreas S. – Tsao, Doris: Catalyzing next-generation Artificial Intelligence through NeuroAI. Nature Communications, 14, 2023. DOI: https://doi.org/10.1038/s41467-023-37180-x
von Neumann, John: The Computer and the Brain, New Haven: Yale University Press, 1958. pp. 1-97.
von Neumann, John: First draft of a report on the EDVAC. Moore School of Electrical Engineering, University of Pennsylvania, 1945. 10.5479/sil.538961.39088011475779 [29 September 2025]

The Turing test appeared and gained popularity in the early 1950s. In connection with the question posed by Alan Turing, 'Can machines think?', Turing's imitation game was intended to demonstrate whether a 'machine' is capable of displaying human-like intelligence. The essence of this test is whether the 'machine' is able to give the impression that the conversation is taking place with a human being rather than a machine.[14] Among the early developments in AI, the emergence of the term 'AI'[15] is noteworthy, which is associated with John McCarthy et al.'s 1956 summer research project at Dartmouth College.

In the 1970s and 1980s, following initial successes, sharp criticism emerged regarding the (future) effectiveness of AI research, which in many cases led to a reduction or termination of funding.

From the 1990s to the 2010s, there were significant breakthroughs in the fields of machine learning and deep learning research. The research results served as evidence for the diverse applicability of AI.

Despite its 'shortness', the period beginning in 2020 and continuing today has brought numerous new technological solutions that are leading to a boom in AI research. Generative AI using large language models and research related to its use are receiving particular attention. Also worth mentioning are content generation (e.g. text, image, sound, video) and AI support, which is often used in searches, as well as the use of AI-based extensions. It can be observed that AI integration is playing an increasingly important role, which is also reflected in the presentation of AI's legal applications specifically for legal work. At the same time, it is important to note that the risks and dangers associated with AI, as well as ethical, sustainability, copyright, regulatory and social issues, have gained prominence during this period and continue to do so today.

[14] Turing, Alan M.: Computing Machinery and Intelligence. Mind, 49., pp. 433- 460. 1950. pp. 433-434.
[15] This is where the term 'artificial intelligence' first appeared.

III. Some aspects of AI regulatory issues

The relationship between technology and regulation often appears to be contradictory.[16] Regulation is a tool that can either hinder or encourage technological change. The relationship depends on the technology of regulation – the formulation of regulatory policy and the choice of tools.[17]

As AI has developed, regulatory, ethical and legal issues have continually arisen, some of which remain unresolved, while different approaches have been taken in different parts of the world with regard to certain issues. At the same time, it is also important to note that general, non-binding principles have been adopted (e.g. the document defining the OECD Principles for AI, which has been updated in 2024[18], and the UNESCO's recommendation adopted in November 2021 focusing on ethical issues and principles of AI[19].

When examining the issue of AI regulation, different positions have emerged worldwide as to whether uniform, comprehensive regulation is necessary, or whether sector-specific legislation for certain activities and the establishment of certain principles are the necessary steps. In addition to the practice applied in the European Union, the AI regulatory practices of the United States of America and China, which are of particular importance from an international perspective, are briefly presented in a comparative manner in the table below, covering the period until the entry into force of the AI Act, without claiming to be exhaustive.

[16] Wiener, Jonathan B.: The regulation of technology, and the technology of regulation. Technology in society 26, 2004., p. 483.
https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1960&context=faculty_scholarship [2025. 10. 12.]
[17] Wiener, Jonathan B., 2004. p. 483.
[18] OECD.AI Policy Observatory: OECD AI Principles overview, Recommendation of the Council on Artificial Intelligence, 3 May 2024, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [12 May 2025.]
[19] UNESCO: Recommendation on the Ethics of Artificial Intelligence, adopted on 23 November 2021, published in 2022, https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi [12 May 2025.]

Criteria	European Union (EU)	United States of America (USA)	China
Regulatory approach	Comprehensive, risk-based approach, uniform legal framework in the form of regulations.	Decentralised, case-by-case approach, characterised by sector-specific legislation.[20]	No comprehensive regulation, characterised by a vertical, technology-specific framework.[21]
Main regulatory instrument	AI regulation and ethical guidelines [22]	The US applies existing federal and state laws.[23] US states are leading the way in AI legislation, while progress at federal level is slow.[24]	Provisional measures for the governance of generative AI services.[25]
Supervisory/regulatory structure	European AI Office[26]	Highly fragmented among federal agencies.[27]	Government oversight.[28]
Risk categories/ risk management	Four-level risk classification (unacceptable, high, low, minimal or none)	The US approach to risk management is generally risk-based and sector-specific.[29]	Prohibited practices, characteristics of public opinion, social mobilisation capabilities.[30]
Effective date – introduction	Entered into force on 1 August 2024. Later entry into force dates have been set for the application of certain provisions.	–	It entered into force on 15 August 2023.

Table 1 – Regulatory approach to AI, practices applied in the European Union, the United States and China

Source: Own compilation based on the references indicated.

[20] Pernot-Leplay, Emmanuel: The AI Dilemma: AI Regulation in China, EU & the U.S., utoljára frissítve: 2024 november, https://pernot-leplay.com/ai-regulation-china-eu-us-comparison/#toc_The_US_Guidelines_and_Narrow_Bills [10 May 2025]

[21] Sajduk, Błażej – Dziwisz, Dominika: Comparative Analysis of AI Development Strategies: A Study of China's Ambitions and the EU's Regulatory Framework, EuroHub4Sino - European Hub for Contemporary China, 20 September 2024, https://eh4s.eu/publication/comparative-analysis-of-ai-development-strategies-a-study-of-chinas-ambitions-and-the-e-us-regulatory-framework [10 May 2025]

[22] AI Act Preamble (27)
It should be noted that other technical materials have also been published to facilitate the practical application of the AI Act, and implementing regulations are expected.

[23] Markevich, Gleb: AI Regulation: A Comparative Analysis of Approaches in the US, EU, and China, 14 July 2023, https://www.linkedin.com/pulse/ai-regulation-comparative-analysis-approaches-us-eu-china-markevich/ [30 April 2025]

[24] Maslej, Nestor – Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Tobi Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, Sukrut Oak. "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, 2025. április, pp. 1-456., p. 21., https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf [21 May 2025]

[25] BakerMcKenzie: China: New interim measures to regulate generative AI, Client Alert, 2023 augusztus, p. 1. https://insightplus.bakermckenzie.com/bm/attachment_dw.action?attkey=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQJsWJiCH2WAWuU9AaVDeFglGa5oQkOMGl&nav=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQbuwypnpZjc4%3D&attdocparam=pB7HEsg%2FZ312Bk8OIuOIH1c%2BY4beLEAezirm3%2BK7wMU%3D&fromContentView=1 [10 May 2025]

Among the latest developments, it is worth mentioning that on 1 August 2025, the Commission and the European Artificial Intelligence Board confirmed that The General-Purpose AI Code of Practice developed by independent experts is an appropriate voluntary tool for providers of general-purpose AI systems to demonstrate compliance with the AI Act.[31] On 19 November 2025, the Commission announced the Digital Omnibus on AI Regulation Proposal, which proposes targeted simplification measures for certain provisions of the AI Act to ensure timely, smooth and proportionate implementation.[32] Months of consultation with industry, small and medium-sized enterprises, civil society and Member States have highlighted that organisations face significant uncertainty regarding the AI Act: the designation of national authorities has been delayed, and shortcomings in harmonised standards and the complex interactions between the AI Act and other EU digital legislation have become apparent. The Digital Omnibus on AI directly addresses these shortcomings.[33] As a result, the AI Act may be amended.

Following the entry into force of the AI Act, various legal norms and legal acts relating to its implementation in Hungary have been continuously published, and the national AI strategy has also been renewed.

In September 2025, the renewed national AI strategy was published[34], which aims to reflect the experience gained during the implementation of the first national strategy published in May 2020[35] and the changes in the dynamic development of technology. The AI strategy, which is reviewed annually, provides a comprehensive overview of the development, application and regulation of AI in Hungary until 2030, sets priorities, particularly in education, research, the economy and public administration, and defines the

[26] European AI Office: https://digital-strategy.ec.europa.eu/en/policies/ai-office [2025. 05. 10.]
[27] Engler, Alex: The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment, 2023. április 25., https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/ [2025. 05. 10.]
[28] Sajduk, Błażej et al.: 2024. szeptember 20.
[29] Engler, Alex: 2023. április 25.
[30] Dorwart, Hunter – Qu, Harry – Bräutigam, Tobias – Gong, James: Preparing for compliance: Key differences between EU, Chinese AI regulations, IAPP, 2025. február 5., https://iapp.org/news/a/preparing-for-compliance-key-differences-between-eu-chinese-ai-regulations [2025. 05. 10.]
[31] European Commission: Commission Opinion on the assessment of the General-Purpose AI Code of Practice, 2025. augusztus 1., https://digital-strategy.ec.europa.eu/en/library/commission-opinion-assessment-general-purpose-ai-code-practice [2025. 11. 05.]
EUROPEAN COMMISSION: COMMISSION OPINION of 1.8.2025 on the assessment of the General-Purpose AI Code of Practice within the meaning of Article 56 of Regulation (EU) 2024/1689, Brussels, 1.8.2025, C(2025) 5361 final
[32] European Commission: Digital Omnibus on AI Regulation Proposal, 2025. november 19., https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal [2025. 11. 20.]
EUROPEAN COMMISSION Brussels, 19.11.2025 COM(2025) 836 final 2025/0359 (COD) Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI)
[33] CMS LawNow™: Digital omnibus on AI: The European Commission unveils a streamlined and more coherent approach to AI regulation, 2025. november 19. (https://cms-lawnow.com/en/ealerts/2025/11/digital-omnibus-on-ai-the-european-commission-unveils-a-streamlined-and-more-coherent-approach-to-ai-regulation [2025. 11. 20.]
[34] AI Strategy, 3 September 2025, pp. 1-119.
[35] Hungary's Artificial Intelligence Strategy 2020-2030: May 2020, Digitális Jólét Nonprofit Kft., pp. 1-58. https://cdn.kormany.hu/uploads/document/6/67/676/676186555d8df2b1408982bb6ce81c643d5fa4ab.pdf [19 October 2024]

main directions for the development of the data economy.[36] The AI strategy specifically mentions the development of AI literacy and critical thinking.

[36] AI Strategy, 3 September 2025, pp. 1-119.

IV. AI literacy and related literacies

IV.1. AI literacy in th AI Act

With regard to EU regulation of AI, it should be noted that it plays a significant role and sets out specific obligations in relation to AI literacy. Article 3(56) of the AI Act defines AI literacy, which means skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause. Article 4 of the AI Act sets out the obligations relating to AI literacy on the part of providers and users. In my point of view, there are a number of factors that influence the assessment and determination of an appropriate level of AI literacy in the context of the above regulation. These include:

- the division of responsibilities, the possibility of supervision as a 'higher-level' control (e.g. attorney-at-law – trainee lawyer in a law firm);
- previous professional experience, participation in training courses, qualifications;
- the stage at which the individual is involved in the use of AI (e.g. in the training process in the case of machine learning or 'running a query at the touch of a button' – the degree of involvement varies);
- how many people are affected by the use of AI to perform the task in question, does it significantly affect the work of others;
- the consequences of a possible error, whether it is possible to correct the error directly;
- what authorisation the person in question has in relation to the AI system in question.

It should be noted that the Digital Omnibus proposal also contains provisions on AI literacy. The Commission proposes to remove the obligations on service providers and deployers regarding AI literacy. Instead of service providers and users being legally obliged to ensure that their employees operating and using AI systems have an adequate level of AI literacy, the Commission and Member States will be required to promote the acquisition of AI literacy and and '*encourage providers and users of AI systems to take measures to ensure an adequate level of AI literacy*'.[37]

IV.2. The scientific approach to AI literacy

According to Duri Long and Brian Magerko, AI literacy is a set of competencies that enable individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use artificial intelligence as a tool online, at home, and at work.[38]

[37] Digital Omnibus on AI

[38] Long, Duri – Magerko, Brian: 2020. What is AI Literacy? Competencies and Design Considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, pp. 1–16., p. 8. DOI: https://doi.org/10.1145/3313831.3376727

The above statements should also be interpreted in relation to 'prompting' in the context of generative AI. Defining the prompt and supplementing it with additional elements and questions plays a key role in collaborating with AI.

Marc Pinski and Alexander Benlian defined AI literacy as the socio-technical competence of people consisting of knowledge and experience, which together form two separate types of competence that constitute AI literacy.[39] In Teresa Heyder and Oliver Posegga's approach, AI literacy consists of functional, critical and socio-cultural dimensions. They identified possible connections between these dimensions and concluded that socio-cultural AI literacy plays a significant role in enabling employees to collaborate effectively with AI in the workplace.[40] In addition to facilitating collaboration, the critical dimension was also identified in this case.

Davy Tsz Kit Ng et al. took an in-depth look at the conceptual approach to AI literacy and its individual dimensions. Based on their research – a literature review of 30 existing, peer-reviewed articles – they proposed four aspects (i.e. knowledge and understanding; use and application; evaluation and creation; and ethical issues) for promoting AI literacy based on the adaptation of classic literacies.[41]

AI literacy	Definition
Knowledge and understanding of AI	Know the basic functions of AI and how to use AI applications.
Use and application of AI	Apply knowledge, concepts and applications of AI in different scenarios.
Evaluation and creation of AI	Higher-order thinking skills (e.g. evaluation, estimation, prediction, planning) with AI applications.
AI ethics	Human-centred considerations (e.g. fairness, accountability, transparency, ethics, security).

Table 2 – Dimensions of AI literacy and their definitions
Compiled by the author based on a study by Davy Tsz Kit Ng et al.

In relation to AI literacy, it can be said that it is closely related to digital literacy, privacy literacy and algorithmic literacy. Beyond the literature, it is important to emphasise that the development of AI literacy is also relevant under the AI Regulation. It should also be noted that critical thinking and critical evaluation are also skills related to AI literacy.

[39] Pinski, Marc – Benlian, Alexander: 'AI Literacy - Towards Measuring Human Competency in Artificial Intelligence', 2023, Hawaii International Conference on System Sciences 2023 (HICSS-56). 3. https://aisel.aisnet.org/hicss-56/cl/ai_and_future_work/3 pp. 165-174; p. 165.
[40] Heyder, Teresa – Posegga, Oliver: 'Extending the foundations of AI literacy', 2021. ICIS 2021 Proceedings. 9. https://aisel.aisnet.org/icis2021/is_future_work/is_future_work/9
[41] Ng, Davy Tsz Kit – Leung, Jac Ka Lok – Chu, Samuel Kai Wah – Qiao, Maggie Shen: Conceptualising AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, Volume 2., 2021., DOI: https://doi.org/10.1016/j.caeai.2021.100041 (Hereinafter: Ng D. T. K. et al., 2021a)

In my opinion, with the rise of AI and the emergence of various AI-based solutions, it is particularly important to develop AI literacy, especially in relation to the limitations of easily accessible generative AI systems. In this area, in addition to professional knowledge, a practical approach plays a key role.

V. Current trends in the use of AI in the legal work

This part presents the results of a total of 12 surveys and reports in detail with the aim of providing a comprehensive overview of current trends in the use of AI and generative AI in legal work from a practical perspective. In the case of all surveys and reports, it is important to note that due to the dynamic development of AI and the significant impact of AI and generative AI on the legal profession and legal work, individual results may change rapidly, even from year to year. At the same time, the trends are evident in the results, both for the present and for the future. The situation of domestic law firms and legal departments is very different from that in the US (e.g. size of law firms, number of employees), but on the one hand, international trends have a significant impact on the domestic legal market, and on the other hand, the use of AI among legal professionals can also be observed in Hungary, so the findings and the effects of AI are relevant.

The main common findings and contradictions identified in these reports and surveys are detailed below.

Common finding:

a) The spread and growth of AI use

The growth in the use of AI has been noted in all reports and surveys. There is consensus that AI has a significant impact on both the legal profession and legal work. Its use in the legal sector is also on the rise. The use of this technology is also growing in individual and small law firms. A high proportion of respondents reported using it at least once a week. The majority of respondents expect the use of AI to become even more widespread in the near future.

b) Types of AI tools

Based on the results, ChatGPT dominated with a high usage rate. It is important to note that this technological solution belongs to the category of general-purpose generative AI systems. Legal professionals use publicly available AI tools at a higher rate than law-specific AI systems. At the same time, it is also important to emphasise that the use of legal-specific AI tools also appeared in the responses, with specific applications being named in some cases. Legal-specific AI tools are primarily used in larger law firms, while smaller and individual law firms tend to prefer free, general-purpose tools. In my opinion, the cost of investing in AI systems may also be a deterrent in this case, given the uncertainty of return on investment.

c) Advantages and effects of using AI

In light of the results, the main advantages of AI were primarily identified as time savings and increased efficiency. In addition, productivity, financial return, cost savings and revenue growth were also mentioned. In this regard, in my point of view the level of development of the organisation prior to the introduction of AI and the development steps it has taken are also decisive factors.

The vast majority of legal professionals who responded agree that AI will have a transformative effect on legal work and that this technology can support their

professional activities. Other effects of AI include changes in billing practices, the inevitable transformation of these practices, and the need to acquire new skills.

d) Areas of legal-specific application for AI

Among the specific areas of application of AI, legal research ranked highest, followed by general research. Another prominent area, expressed as a collective term, is document management, which broadly covers document review, summarisation, analysis and data extraction. In addition to these, text composition, letter (email) preparation, translation, case strategy and e-discovery also appeared.

e) Factors influencing the use of AI

In addition to technological conditions, the size of the law firm and the introduction of AI and its integration into work processes can be identified as influencing factors. It is important to emphasise that the knowledge and proficiency of legal professionals in AI is also of decisive importance, as is whether the organisation in question has the possibility of using a legal-specific AI system and whether it has AI regulations in place that define the framework for its use. With regard to AI regulations, it should be noted that the results showed that, regardless of the type of organisation, a significant proportion of law firms, legal departments and companies do not have such regulations in place.

f) The role of training and knowledge expansion

There was general consensus among legal professionals on the importance of training, with mandatory training appearing in some cases. Among the preferred learning methods, a practical approach was emphasised, on the one hand through experimentation with AI tools and, on the other hand, through reading written materials in order to learn about current developments. The results also draw attention to the fact that although legal professionals have AI knowledge in many cases, they consider their knowledge of practical application to be insufficient.

Contradictions and differences:

a) The usability and usefulness of AI and the regularity of its use

The results showed that the vast majority of legal professionals see the potential of AI, i.e. they consider it ‘can use’, but do not consider it ‘should use’. A contradiction can be observed between the theoretical and practical approaches. Furthermore, the regularity of use also raises questions, as in many cases initial experimentation is not followed by regular use.

b) Trends in the use of legal-specific AI

The latest report (Clio) shows a significant decline in the use of law-specific AI systems, while other surveys have found an upward trend. A slight decline can also be observed in some law-specific areas of application.

c) Law firms and legal departments

Due to the specific nature of the professions, there are some differences in the results in terms of the frequency of AI use, the existence of AI regulations and efficiency gains. However, differences in survey methodologies do not allow for clear conclusions to be drawn and differences to be identified.

d) Availability of training

There were also contradictions between some surveys regarding the availability of training. In some surveys, the majority of respondents stated that AI training was not available to them, whereas in another survey, it was found that the majority of respondents were required or would be required to participate in AI training. and there were also cases where the frequency of training was also a factor to be examined. This suggests that there are significant differences between organisations in terms of the availability of training.

e) The use of AI by lawyers in client work

Overall, it can be concluded that there is currently a contradiction, and in my opinion, there is not enough information available to determine whether clients view the use of AI in client work positively or negatively. It should also be noted that practice is ambivalent with regard to transparency concerning the use of AI.

f) Increased use of AI vs. preparedness

The increase in AI knowledge and the relatively rapid spread of AI in legal work is not accompanied by a proportional increase in preparedness.

General comments

In my opinion, there are numerous opportunities for supporting legal work through the use of AI and generative AI, which should be exploited as soon as possible. However, this requires both the necessary material and human resources. Various technology companies are making outstanding efforts to make AI technology tools user-friendly and easy to use, so that – in everyday terms – the technology is available and no high-level technological knowledge, such as software development, is required to use it. At the same time, data security, organisational-level data protection, the integration of AI systems into internal systems, and organisational preparedness present challenges. The importance of this is undeniable, but in my opinion, the more important element is personal conditions. By personnel requirements, I mean that users should have the appropriate knowledge to use AI consciously and responsibly, which also means that they should be aware of the limitations that may apply when using AI and take these into account. Furthermore, continuous training is essential in this dynamically changing field. In my opinion, it is important to have an AI policy in place and to establish a framework for the use of AI, on the one hand to ensure information security and data protection, and on the other hand, to ensure that employees use AI systems approved by the organisation for their work, thereby reducing the risk of employees using AI systems that may pose a risk, i.e. employees know which AI tools they can use ‘legally’ and which they cannot.

With regard to the surveys and reports detailed in this chapter, it is important to note that the results of the surveys should be interpreted with caution. It is likely that a larger number of people who are more interested in and open to technology, and who use AI to a greater extent in their legal work, participated in the surveys. It is also important to note that some surveys were conducted specifically among individuals for whom AI is no longer a novelty and who have a certain level of AI knowledge. Furthermore, it should be mentioned that the authors of some reports are present in the legal market with some kind of AI tool, AI-supported product or service. In my opinion, legal professionals who do not use this technology or use it only to a limited extent in their legal work are underrepresented in these reports and surveys. The opinions and practices of those who do use it are expressed to a greater extent. The positive effect of this is that even those who are not currently at the forefront can see the development opportunities that AI offers. The reports and surveys are not consistent in terms of whether the results and findings refer exclusively to generative AI or AI in general, and the concept of AI has not been explicitly defined, which also allows for discrepancies.

I also consider it important to note that the regulatory system in force in a given country is also a factor influencing the application of AI systems. US regulations on both data protection and AI differ significantly from those applicable in the European Union. With regard to domestic practice, it is important to emphasise that both the legal system and economic and territorial differences have an impact. The structure of the legal profession in Hungary is completely different from that in the US.

The results of these reports and surveys provide insight into both the present and the future, showing what is happening and what legal professionals can expect in relation to the use of generative AI and AI in legal work-

VI. AI-supported applications in legal work

This chapter presents a few international and domestic AI-supported applications that can assist legal professionals in performing ‘classic’ legal tasks. Given that this field is changing extremely dynamically, with new functions and solutions constantly emerging, many companies and start-ups prove to be short-lived in practice, with the company transforming, focusing on developing new products or ceasing to exist, as in the case of ROSS Intelligence[42] . In this case, my aim is not to provide a comprehensive overview of all solutions used in all areas of law, but merely to provide a snapshot of the current possibilities.

It is important to note that not all AI-supported solutions mentioned in this chapter are necessarily applicable in the domestic context, due to linguistic or legal constraints. At the same time, they can serve as guidelines and possible areas of application for the (near) future. It can be observed that software providers are finding solutions to overcome potential language barriers at an increasingly rapid pace. Given that the research focuses primarily on

[42] The Founders: ROSS Intelligence: Announcement, 11 December 2020, https://blog.rossintelligence.com/post/announcement [21 May 2025]
The ROSS platform will no longer be available from 31 January 2021.

law firms and legal departments, the applications also relate to solutions used in these organisations.

As early as 2020, AI was generally used in six ways in the legal field, namely: e-discovery; expertise automation; legal research; document management; analysis and generation of contractual and litigation documents; and predictive analytics.[43] This approach and categorisation can still be considered valid today, although there may be differences due to regional factors, and over time, as AI has developed, a few other areas of application have been added. These include agent-based AI solutions, legal chatbots, and the use of generative AI systems in support of internal processes. It should also be noted that AI-based solutions in the areas of *document review*, document management and processing, *drafting*, contract analysis and review have grown and developed significantly in recent years. It is also important to note that the emergence and spread of generative AI and the continuous development of large language models (LLMs) have led to an expansion of application areas. The role and practical application *of compliance* has also become more important in the life of companies.

The table below summarises the various functions of the AI applications described in this chapter.

Function/ Service provider/ Service	**Legal research**	**Document management**	**MI agent**	**E-discovery**	**Predictive analytics**	**Support for internal work processes**
Bloomberg Law	×				×	
Harvey	×		×			×
LEGALFLY	×	×	×			
LexisNexis	×		×		×	
RobinAI	×	×	×			×
CoCounsel Legal	×					
Wolters Kluwer	×		×			
ORAC Kiadó	×					
Luminance		×		×		×
Relativity		×		×		
AIrite						×

Table 3 – AI-supported applications in legal work
Source: Compiled by the author based on the AI-supported applications examined.

[43] Davis, Anthony E.: The Future of Law Firms (and Lawyers) in the Age of Artificial Intelligence, 2 October 2020, The Professional Lawyer Vol. 27, No. 1. October 2020, https://www.americanbar.org/groups/professional_responsibility/publications/professional_lawyer/27/1/the-future-law-firms-and-lawyers-the-age-artificial-intelligence/ [21 May 2025]

VII. Legal LLM benchmarks

In the context of generative AI systems, the emergence and proliferation of various and increasingly numerous large language models (LLMs) has led to the use of benchmarks to compare the performance and functioning of these models. Numerous comparison methods and models allow LLMs to be compared according to various criteria.

Equipping LLMs with legal expertise can not only improve the efficiency of legal professionals' work, but also meet the overwhelming demand for legal assistance from non-professionals, thereby improving public access to justice.[44] Based on the results of the survey described above, it has been established that many users also use generative AI systems in connection with legal issues. In this regard, it should be noted that a critical approach to conscious use is also essential when used by non-legal professionals. At the same time, LLM solutions based on pre-checked, appropriate training data can provide real help to both legal and non-legal professionals.

VII.1. General evaluation metrics

General evaluation metrics include accuracy metrics: accuracy, F1-score and exact match; text-level comparison metrics include ROUGE and ROUGE-L; hallucination rate; scoring method; and semantic similarity method.

In summary, it can be concluded that these methods have various limitations. In terms of scoring metrics, there is a subjective element, and in the case of ROUGE, there is word or n-gram overlap, which means that the number of words may match, but the content may be completely incorrect. In my opinion, the primary consideration should be the accuracy of the content, which can be supplemented by a scoring method, accuracy, and linguistic evaluation.

The most recommended approach to content evaluation is to establish a unique set of criteria and key phrases and then evaluate the correctness of the responses either by a scoring method or by a 0-1 evaluation of whether the given content is correct or not.

VII.2. Main findings of the legal LLM benchmarks

The various legal LLM benchmarks reviewed (LegalBench, LawBench, LEXTREME, LexGLUE, SCALE, LBOX OPEN) show that, on the one hand, there is no LLM that performs best in all tasks, and on the other hand, there are limitations to the methods and metrics used, with manual expert evaluation being highlighted.

[44] Cui, Junyun – Shen, Xiaoyu – Nie, Feiping – Wang, Zheng – Wang, Jinglong – Chen, Yulong: A survey on legal judgment prediction: Datasets, metrics, models and challenges. 2022. Preprint, arXiv:2204.04859.
Trozze, Arianna – Davies, Toby – Kleinberg, Bennett: Large language models in cryptocurrency securities cases: can a gpt model meaningfully assist lawyers? Artificial Intelligence and Law, 2024. pp. 1–47.

Legal LLM benchmark/ Criterion	LegalBench	LawBench	LEXTREME	LexGLUE	SCALE	LBOX OPEN
Language	English	Chinese	24 European languages	English	German French Italian Romansh English	Korean
Evaluation method, Measuring instrument	Exact match, F1 score, balanced accuracy, Manual expert evaluation.	Accuracy; F0.5; F1-; rc-F1-; soft-F1-score; n-logarithmic distance, Rouge-L.	Macro-F1 score harmonic mean.	Micro-F1, macro-F1 score.	BERTScore, BLEU, METEOR, ROUGE, hierarchically aggregated macro F harmonic mean, NDCG, capped recall.	Exact match, F1 score, Rouge-1, Rouge-2, Rouge-L.
Examination criteria	Legal reasoning skills. Various types of models.	Memorisation and understanding of legal knowledge understanding and application of legal knowledge.	Multilingualism, law-specific vs. general models.	Law-specific vs. general models.	Domain-specific knowledge, multilingual comprehension, multitasking abilities.	Impact of domain-specific corpus, domain adaptation.
Main findings	GPT-4 outperforms GPT-3.5 and Claude-1.	GPT-4 was unable to accurately reproduce the content of legal texts, with hallucinations being common. Law-specific LLMs do not necessarily perform better than general-purpose, Chinese language-specific LLMs.	Larger models perform better. Supervised approaches outperform ChatGPT.	No single model performs best on all tasks.	The models tested, including ChatGPT, perform poorly, especially in court reasoning generation and information retrieval tasks.	It is important to pre-train language models on large-scale, domain-specific corpora.

Table 4 – The legal LLM benchmarks examined
Source: Compilation based on the benchmarks examined.

The main findings of the empirical studies are that, on the one hand, based on the various models tested, it is not possible to declare a generally best system, as the same model performed better in certain tests and less well in others, and the results of the studies are not consistent. On the other hand, no far-reaching conclusions can be drawn based on the metrics.

Given that my research is linked to amendments to specific legal provisions, the methods presented in this chapter did not prove to be entirely appropriate, particularly in view of their limitations of application, so I developed my own set of analytical criteria for the case study, which will be described in detail in the next chapter.

VIII. Empirical research

Surveys of the legal sector have found that the use of general-purpose generative AI systems is significant among legal professionals in connection with their work, and that non-legal professionals also turn to generative AI systems for legal issues.

The aim of my empirical studies is to use examples to show how effective different generative AI systems are in tracking legislative changes. A further objective is to identify common types of errors in the responses generated by the generative AI systems examined and to examine the developments of the generative AI systems examined in relation to specific responses.

A review of the literature shows that AI systems are increasingly capable of providing support in performing law-specific tasks, and that more and more functions are appearing, showing promising results in various areas of activity. For example, they are used in practice in legal research, certain repetitive tasks, the preparation of various legal documents, the review of contracts and the preparation of summaries.

Nevertheless, with regard to the limitations of generative AI systems, it is worth paying attention to how the system in question relates to the examination of validity and the tracking of legislative changes, as the collection, analysis, processing and display of sources during the final text generation provides information about the functioning of the generative AI system in question. There are also legal search interfaces available, both free of charge and for a fee, for displaying the legislation in force and examining its validity and various time statuses. However, when using different generative AI systems, it is important that the generative AI system in question responds to the question asked with a correct or incorrect reference, taking into account the applicable legal provisions.

VIII.1. Methodological framework

In the case of specific generative AI systems, I used qualitative exploratory research to examine the extent to which these systems are capable of answering different types of simple legal questions in the context of specific legal provisions and of tracking changes in legislation, as well as the types of errors that occur frequently.

The qualitative exploratory research consisted of a preliminary examination and an empirical research, in both of which I examined the responses generated by specific generative AI systems using comparative text analysis.

With regard to the legal LLM benchmarks described and the related metrics, it was found that there are certain limitations to the application of the metrics. Given that my research is related to specific legal provisions and their amendments, the examination of content is of primary relevance. In this case, the grammatical analysis was only relevant in relation to the content, but due to the nature of the legislation, accuracy was of paramount importance. The scoring method was not used in the analysis, as the evaluation of answers to different types of questions could have been subject to subjective elements in terms of the correctness or inaccuracy of the answers. The role of human evaluation was emphasised in the legal LLM benchmarks. Taking into account the limitations of the metrics, I created a

unique set of analysis criteria, which I used to perform the comparative text analysis. During the evaluation, I examined the correctness of the answers generated by generative AI systems in relation to the legal provisions and the time period under review.

During the text analysis, I manually analysed each response to each question based on the analysis criteria. In this case, I present the results not by question, but by identified error type.

The research consists of a preliminary examination conducted in 2024 comprising 33 questions and an empirical data collection conducted in two stages in 2025 comprising 19 questions. One of the purposes of the examination was to examine the effects of developments affecting generative AI systems, so a significant amount of time passed between the two studies, which was appropriate from a research perspective, given the dynamic development of AI systems.

The dissertation highlights some of the results, focusing on whether the errors identified in the preliminary examination may occur in the empirical research conducted in 2025, as well as what differences can be detected through the development of the models, and whether any recurring, "unresolved" errors are visible. It is important to note that, based on this case study, no far-reaching, clear conclusions can be drawn about the functioning of the generative AI systems examined, but they can serve as examples and guidelines for the use of generative AI systems, the identification of errors, the consideration of limitations, and the conduct of further research

Labour law legal institutions and related legal provisions are presented only to the extent necessary for the analysis of responses generated by generative AI systems, in accordance with the criteria detailed below.

The summary table below provides an overview based on the main aspects of the preliminary examination in 2024 and the empirical research in 2025.

Criterion	Preliminary examination	Empirical research
Purpose	Examination of legislative amendments affecting specific provisions of the Labour Code, identification of possible error types in the MI systems examined, and formulation of findings.	Monitoring legislative amendments affecting paternity leave in the Labour Code, with particular regard to the latest amendment effective from 1 January 2025, and re-examining the possible occurrence of error types identified in the preliminary examination by means of new queries. Examining the possible occurrence of error types identified in the preliminary examination by means of repeated queries, identifying recurring errors that "still appear". monitoring the development of the generative AI systems examined in relation to the present topic, and reviewing and supplementing the

Criterion	Preliminary examination	Empirical research
		findings of the preliminary examionation.
Testing tool	A preliminary, systematically compiled questionnaire related to the topic of paternity leave.	A narrowed-down version of the questionnaire used in the preliminary examination, modified and supplemented in certain respects, relating to the topic of paternity leave (parental leave is not part of the research).
Method	Comparative text analysis according to specific examination criteria.	Comparative text analysis expanded with new criteria.
Number of questions in the questionnaire	A total of 33 questions.	A total of 19 questions, of which three are new and belong to the introductory questions.
Feedback, interaction, further clarifying and supplementary questions	No feedback was provided on the responses generated by the AI systems examined. No clarifying questions were asked in the event of an inappropriate response.	There was no feedback on the responses generated by the AI systems examined. Some supplementary questions and prompts (e.g. source) appeared systematically among the questions in the questionnaire during the query and 'conversation' with the AI system.
Time of the query	The questions were asked on 6 and 7 August 2024.	The questions were asked twice, on 18 and 19 May 2025, with regard to certain additional questions. On both occasions, the entire questionnaire was the subject of the query.
Generative AI systems used	A total of eight generative AI systems: - Anthropic – Claude 3.5 Sonnet - Anthropic – Claude 3 Opus - Anthropic – Claude 3 Haiku - Google – Gemini - Google – Gemini Advanced - Open AI – ChatGPT 3.5 - Open AI – GPT-4 - Open AI – GPT-4o	A total of four generative AI systems: - Anthropic – Claude 3.7 Sonnet - Google – Gemini 2.0 Flash - Open AI – ChatGPT-4o - Open AI – GPT-4.1
The "location" of the query	The query was performed on the web interface of ChatGPT 3.5, GPT-4 and GPT-4o via the "Chatbot App"[45] ; on Claude 3.5 Sonnet, Claude 3 Opus and Claude 3 Haiku via Claude[46] ; and on Google Gemini and Gemini Advanced via Gemini[47] .	The query was performed on ChatGPT-4o and GPT-4.1 via the ChatGPT web interface[48] ; on Claude 3.7 Sonnet via the Claude web interface[49] ; and on Gemini 2.0 Flash via the Gemini web interface[50]

[45] ChatbotApp: https://chatbotapp.ai/ [06-07 August 2024]
[46] Claude: https://claude.ai/ [06. August 2024]
[47] Gemini: https://gemini.google.com/ [07 August 2024]
[48] ChatGPT: https://chatgpt.com/ [18-19 May 2025]
[49] Claude: https://claude.ai/ [18-19 May 2025]
[50] Gemini: https://gemini.google.com/ [18-19 May 2025]

Criterion	Preliminary examination	Empirical research
Primary legal provisions examined	Labour Code Section 118(4) Labour Code Section 122(4a) Labour Code Section 118/A(1) Labour Code Section 127(1)	Labour Code Section 118(4)
Frequently occurring error types	- Terminology error. - Determining the length of paternity leave. - In cases establishing entitlement to paternity leave, completeness was not ensured.	- Use of inappropriate websites for responding. - Contradictions between responses generated by a given generative AI system. - Incoherent responses. - Inappropriate legal references.
Limitations encountered during the query	Time limit: After a certain number of questions had been run, a "waiting time" was applied, which in the case of Claude was five hours, and, in the case of the Chatbot App, due to system overload, the query option was unavailable for 24 hours.	Quantitative limitation: In the case of Claude, the text "maximum conversation length reached" appeared during the query on 19 May 2025, making it necessary to open a new context window.

Table 5 – Comparison of the preliminary examination of the case study in 2024 and the empirical research in 2025

When compiling the questions, I primarily examined the legal provisions of the Labour Code listed above, however, the responses generated by the models of the generative AI systems examined also contained text passages and legal provisions that justified the secondary analysis of additional related legal provisions in the preliminary examination and empirical research.

VIII.2. The uniqueness of the exploratory research

In my research, I examined the capabilities of generative AI systems in a professional-specific field using a new approach. The significance of this lies in the fact that in order to answer a legal question, it is essential that the answer be given in accordance with the current state of the law, i.e. the system must identify the applicable legal provision. Furthermore, if a law has been amended, the generative AI system must take this factor into account when providing an answer.

No research results of this or a similar nature have been published in Hungary. As mentioned in the legal LLM benchmarks, English and English-language data sets dominate the world of generative AI systems. A unique feature of my research is that the queries (questions) were made in Hungarian in relation to Hungarian legal provisions.

According to my own set of analytical criteria, the text analysis was performed using the testing tool (questionnaire) I developed. It should also be noted that my research reflects a dynamically changing environment, taking into account the rapid changes in generative AI systems. Multiple queries made it possible to develop the test criteria and questionnaire ‘on the fly’ in response to changes. In addition, it should be emphasised that the development of generative AI systems was also included as an analytical criterion through multiple queries.

VIII.3. Preliminary examination

VIII.3.1. Methodology of the preliminary examination

The survey conducted in August 2024 can be considered the precursor and basis of the empirical research. Taking into account the results of this research, it can be stated that, when analysing the responses generated by the generative AI systems under investigation to the questions asked in the survey, a number of errors were identified. The errors that occurred are presented systematically, classified into categories with examples. The errors identified during the preliminary investigation served as the basis for the empirical research conducted in 2025.

Methodology of the preliminary examination

The preliminary examination conducted in 2024 served as a precursor to the empirical research that forms the central part of this case study. During this examination, eight large language models of the generative AI system, capable of generating different types of text, were examined. AI systems capable of text generation[51] were subjected to comparative text analysis with regard to specific labour law institutions and related legal provisions. A pre-compiled, systematic set of questions served as the investigative tool for the preliminary investigation. The questionnaire contained a total of 33 questions (prompts) related to the topic of leave in connection with paternity leave as defined in Section 118(4) of the Labour Code and the granting of leave under Section 122(4a) of the Labour Code. In addition, some questions also referred to parental leave [Section 118/A(1) of the Labour Code][52] and maternity leave [Section 127(1) of the Labour Code][53].

The subject of the examination – the legal provisions specified in the Labour Code

The question arises as to why the above-mentioned legal provisions were the subject of the preliminary investigation and why certain legal provisions subsequently served as the basis for empirical research in a modified, narrowed form of the questionnaire. The general selection criteria and the specific selection criteria for each individual legal provision are detailed below.

In general terms, the selection criteria include the fact that the field of labour law affects a wide range of people, that the current labour code came into force at least ten years ago, according to the official gazette, and that it has been comprehensively amended several times since then. The purpose of this preliminary examination was to assess the tracking of current changes and amendments to the legal provisions specified in the selected topic by the generative AI systems under review, with a view to identifying frequently occurring errors.

[51] The eight generative AI systems that form the basis of the testing tool will be described in detail later.
[52] Section 118/A(1) of the Labour Code: *'An employee is entitled to forty-four working days of parental leave until their child reaches the age of three.'*
[53] Section 127(1) of the Labour Code: *'The mother is entitled to 24 consecutive weeks of maternity leave, of which she is required to take two weeks.'*

The specific reasons for selection are described below, in line with the individual legal provisions.

Paternity leave [Section 118(4) of the Labour Code]

In particular, the selection of this legal institution from among the legal provisions is justified by the fact that Section 118(4) of the Labour Code was significantly amended with effect from 1 January 2023.

From the point of view of textual analysis, it is relevant that this legal provision includes the following amendments that have entered into force:

- The term *'additional leave'* no longer appears in the legal provision under review; the legal institution has been defined as paternity leave (terminology).
- This legal provision specifies the cases in which the father is entitled to paternity leave: at the birth of a child, at the adoption of a child, and if the child is stillborn or dies. As a result of the amendment to the law, the cases giving rise to entitlement to paternity leave (hereinafter: cases or cases giving rise to entitlement to paternity leave) have been extended to include the case of *'adoption of a child'* (new element).
- The length of paternity leave has been increased. Under the previous regulations, fathers were entitled to five working days of additional leave in the event of the birth of a child and seven working days in the event of the birth of twins. Under the current legal provisions, paternity leave is ten working days. Another difference is that the Labour Code no longer contains an explicit provision on the length of leave in the case of the birth of twins (different regulation, modified digits written in letters).
- A new element in the rules governing the granting of paternity leave is that the leave *'shall be granted in no more than two instalments'*[54] (new element).

It is important to note that, with regard to this legal provision, the relative dispositive rule of the Labour Code applies, according to which it is possible to deviate from this provision in favour of the employee. Due to the complexity of the regulation, the practical application of this rule on relative dispositivity was not specifically included in the questionnaire and did not form part of the analysis

The analytical criteria of the preliminary examination

In addition to the examination criteria related to the content of the legal provisions detailed above, the examination criteria for the responses generated by the generative AI systems examined included the following: the correctness, completeness, structure and organisation of the responses; the types of errors that may occur and their possible systematic occurrence; the significance of the question and command definition ("prompting"), and the possible occurrence of hallucinations. The questions did not contain any additional information on the textual context in which the question should be "interpreted"; for example, the generative AI system was not required to assume the "personality and expertise" of a lawyer when responding, nor were any length restrictions imposed on the

[54] Labour Code Section 118(4)

responses. The latter was particularly relevant in terms of getting to the point and avoiding text generation that was irrelevant to answering the question, i.e. "irrelevant" text generation.

Given that this study was conducted as a preliminary examination, the results are presented from a systematic analytical perspective, focusing on those questions and answers that are relevant to empirical research. For this reason, not all question-answer analyses are presented.

From a technical point of view, the type of question was significant, i.e. whether it was a closed question or an open question containing additional elements such as "in Hungary" or "the current Labour Code". Furthermore, the structure and organisation of the answers; the clarification of technical terms, i.e. whether the question refers to additional leave to which the father is entitled or paternity leave. It was determined as a technical aspect to be examined how the special characters and numbers in the question affect the answer as a whole. Given that time is still extremely important today, it is particularly important to consider how relevant a response is and whether it contains any irrelevant content. In connection with text generation, another technical aspect to mention is hallucination, both in terms of its existence and its frequency of occurrence.

From a legal perspective, the appropriate and consistent use of technical terms and terminology plays a key role in determining the criteria for analysis. The word "and" in the previous sentence should be interpreted as a conjunctive condition, since from a practical point of view, it is not the single correct answer that is truly relevant, but rather the consistent appearance of answers with the correct content during text generation. The legal relevance of answering the question defined as the objective of the research, i.e. examining the ability of the generative AI models under investigation to follow legislative changes, is beyond doubt. A related factor from a legal perspective is the existence and adequacy of legal references, which includes both the specific legislation and the location of the legislation. From a legal perspective, a distinction can be made between questions that are formulated in general terms, such as those relating to legal institutions, and specific questions, such as those aimed at determining a specific time period.

VIII.3.2. Some results of the preliminary examination

The above-mentioned questions and topics were discussed in detail, given that the main types of errors became apparent from the answers to these questions, and similar errors occurred in the other questions and answers.

Numerous terminological errors appeared in connection with the term ‘paternity leave’ and, in general, with the legal text contained in Section 118(4) of the Labour Code; for example, frequent terminological errors included the inappropriate and inconsistent use of the terms ‘day’ and ‘working day’. It should be noted that if the question deliberately contained inappropriate terminology, the generative AI systems examined used inappropriate terminology. Within the answers and during the examination of successive answers, frequent contradictory answers and contradictory content were found. This led to the conclusion that the answers were inconsistent. The extent of paternity leave was not consistently and adequately indicated in many cases, and the precise determination of the extent of paternity

leave available in the case of the birth of twins posed an even greater challenge for the models examined. It can be observed that when referring to the text of the law, not all essential response elements were displayed, despite the misleading use of quotation marks to indicate the reference to the law. Not only did the reference to the legal provision fail to meet the accuracy criterion, but the definition of the legal provision was also inaccurate in many cases. In the responses, it was common to see content that was irrelevant to the questions being answered. In many cases, the criterion of completeness was not met when it came to listing the circumstances that justify taking paternity leave; typically, the circumstance of child adoption was not listed. There were also examples of hallucinations among the responses, both in relation to paternity leave and the presentation of the text of the legal provision. Incorrect indication of the date of entry into force was also among the types of errors identified. With regard to the use of sources, it is not known on what basis the responses are made. Where multiple reference links and sources were indicated, it was not clear what selection method was used to display a given source, nor exactly what factors were decisive in the selection and display of given sources. It became apparent that tracking legislative changes caused difficulties for the generative AI systems examined, as a significant proportion of the changes to the legislative provisions examined were based on previous legislative provisions that were no longer in force at the time of the query, which should be considered a significant error.

VIII.4. Empirical research

VIII.4.1. Reasons for conducting empirical research

The preliminary investigation and the experience gained from the continuous use of generative AI systems played a key role in the modification of the questionnaire used as the basis for the preliminary investigation. Furthermore, the present empirical research made it possible to examine the frequently occurring error types identified in the preliminary examination from other perspectives, namely what impact the various developments that occurred between the aforementioned studies had on the identified error types, how the models have evolved, and what new error types may have emerged

VIII.4.2. Methodology

The questionnaire forming the basis of the empirical research was administered on two occasions, on 18 May 2025 and 19 May 2025, given that after the first survey, a new analytical perspective arose in connection with the referenced sources through a supplementary prompt. This aspect was the examination of compliance with the instructions specified in the prompt regarding the legal search interfaces, which will be described in detail later.

Unlike the preliminary examination questionnaire, in these cases, emphasis was placed on interaction with the models of the generative AI systems under examination, and additional questions and prompts were included. It is important to note that the purpose of these additional questions and prompts was to gather additional information, particularly with regard to sources, and to gain a deeper understanding of certain aspects of the functioning of generative AI systems ( ). With the exception of the insertion of the above-

mentioned additional questions and prompts, the set of questions was queried in the same order in both cases, in Hungarian.

During the query on 18 May 2025, none of the generative AI system models examined displayed any information indicating that a new "conversation" would need to be started due to the context window reaching its maximum size. In the case of the query on 19 May 2025, the information displayed for the ninth question in Claude 3.7 Sonnet made it necessary to start a new "conversation" from that question onwards, presumably because the maximum context window size had been reached; in other cases, no information message with similar content appeared. The following chapter details the questionnaire used as the basis for the empirical research.

VIII.4.3. The questionnaire used as the basis for the empirical research

Reasons for modifying the questionnaire

During the preliminary examination, when analysing the responses generated by the generative AI systems under investigation, it was established in connection with the identified error types that examining the questions related to Section 118(4) of the Labour Code would be sufficient to achieve the objective of the empirical research. Given that the error types identified in the responses to the 33 questions in the preliminary examination recurred in various legal provisions, questions relating to maternity leave and parental leave were not included in this questionnaire.

The main objective of the empirical research is to examine the compliance of the generative MI systems under review with the legislative amendments to the Labour Code concerning paternity leave, with particular regard to the latest amendment, which will enter into force on 1 January 2025. In addition, further objectives include examining the possible occurrence of error types identified in the preliminary examination through new queries, identifying recurring errors that "still appear", monitoring the development of the generative AI systems examined in relation to the present topic, and reviewing and supplementing the findings of the preliminary examination.

In order to achieve the above-mentioned objectives and taking into account the results and experiences of the preliminary examination, the questionnaire used as the basis for the preliminary examination was reviewed, resulting modifications in terms of both structure and content.

VIII.4.4. Analytical aspects of empirical research

As described in the previous subchapter, the types of errors identified during the preliminary examination had a significant impact on the compilation of the questionnaire underlying the empirical research, and they also play a prominent role in the analysis criteria.

The analysis focused on examining whether the frequently occurring error types identified in the preliminary examination occur in the responses generated by the generative AI systems examined in the empirical research, and whether new error types appear, in connection with monitoring legislative changes. Based on this driving force and taking into account the analytical aspects of the preliminary examination, the following aspects were defined in the empirical research with regard to error types:

- the appropriateness and consistency of terminology use;
- appropriateness and consistency of references to legislation;
- monitoring, appropriateness and consistency of legislative amendments in terms of content and timing (date of entry into force);
- appropriateness and consistency of values related to measurement;
- completeness and consistency in relation to case studies;
- the existence and frequency of hallucinations;
- the role of question formulation (decision-making question or open question with additional information);
- the role of inserted prompts;
- the response used in the case of questions containing incorrect expressions – possible correction;
- the appropriateness and professional nature of the content and form of sources and reference links;
- conciseness, possible presence of irrelevant content that is not necessary to answer the question.[55]

Considering the differences between the queries of 18 May 2025 and 19 May 2025, and taking into account the limitations, the comparison of certain responses also formed part of the analysis.

VIII.4.5. Generative AI systems used in empirical research

In the preliminary examination in August 2024, a total of eight responses generated by generative AI systems associated with three development companies were examined. In the period between the preliminary examination and the empirical research, a number of innovations and new models appeared in generative AI systems as a result of developments. Taking into account the emergence of new models and the results of the preliminary examination, it became necessary and appropriate to review, modify and narrow down not only the set of questions but also the generative AI systems under investigation.

Reasons and criteria for selection

Among the leading developers of generative AI systems, the general-purpose solutions from Anthropic, Google and OpenAI were selected, as they were among the newest and "smartest" models at the time of the queries. In addition, the ChatGPT-4o model was also used, given that, despite the time that had elapsed, it was still listed among the GPT models with outstanding results at the time of the query.

For all models examined, queries were made via subscriber accounts on their own websites[56] . In this case, the comparison of responses generated by the free versions of and

[55] In this context, it is important to note that the number of output tokens differs in the generative AI systems examine d, which may also be a contributing factor.

[56] Claude 3.7. Sonnet: https://claude.ai/new; Gemini 2.0 Flash: https://gemini.google.com/app?hl=hu, ChatGPT-4o and GPT-4.1: https://chatgpt.com/, in all cases: [20 May 2025].

the subscription versions was not a consideration. A significant factor in this is the token limit applied in the free versions.

VIII.4.6. The applicability of the examined generative AI systems in legal work

In light of the results of the preliminary examination and empirical research, the application and operational limitations associated with the examined generative AI systems were identified as common types of errors that must be taken into account during their use. Based on the above, it can be concluded that the generative AI systems examined can be used with caution in legal work, but their use must be conscious and responsible, and it is essential to be aware of the general operating mechanisms, limitations and "potential for error" of these systems.
When responding, it is important to consider how the given generative AI system arrives at the given result, which is relevant from the point of view of other application functions.

It should be noted that the applicability of the models examined in relation to this topic may raise questions. Responding in accordance with the relevant legal provisions is crucial in legal work, and verification is essential. At the same time, the generative AI systems examined can be used to support legal work, taking into account the limitations. It is also important to emphasise that, as a result of the dynamic developments in this field, significant results are emerging in a short period of time.

The results of the research have highlighted that when using the generative AI systems examined, *it is highly recommended to check* the following in the responses.

- The sources and reference links listed by generative AI systems, and in the case of reference links, their time status. When referring to legal texts, if an official, reliable legal search interface appears among the sources, it is necessary to check not only whether the legislation sought appears on that interface, or whether another piece of legislation appears, but also whether the legislation sought appears in its current state, or whether it was placed among the sources according to its previous state.
- The completeness and accuracy of the terminology, dates and time periods appearing in the response.
- In the case of a citation of legislation with a time stamp, the content must be complete and accurate.
- The logical reasoning behind the answers, i.e. it is necessary to check not only the result, but also the logical chain that led to it.

VIII.5. Findings

Based on the results of preliminary investigations and empirical research forming part of the exploratory research, frequently occurring error types and certain application limitations were identified in the generative MI systems examined.

Taking into account all the results of the exploratory research, *there is no "best"* generative MI system among the MI systems examined. It is important to note that the generative MI system that answers the questions most correctly may vary from query to query . At the same time, this leads to another significant challenge for these systems, namely

the illusion of consistent responses. Various examples have shown that contradictory content is common both within responses and between successive responses, which also highlights the accuracy errors of these systems. In the case of these systems, verification is essential. In my opinion, the conscious and responsible use of these technological solutions involves knowledge of the limitations of these systems and the types of errors that frequently occur, i.e. 'knowing what to look out for' when using them. The study focused primarily on identifying common errors and the ability to track legislative changes, but the findings of the research have broader applications, given that legal provisions are an integral part of legal work. In addition to consistent responses, *misleading effects* were also identified as a risk factor in several cases.

With regard to the management and use of sources, the *'black boksz' effect* prevails, and it is unclear what selection methods and criteria are used to display sources. In the case of the generative AI systems examined, there were instances where a given generative AI system did not follow the instructions, and sources appeared in the responses that did not meet the criteria specified in the prompt for legal search interfaces, and official legal search interfaces were not given priority. Nevertheless, in light of the results, it was found that the selection of the appropriate source is of cardinal importance for the correctness of the response. Not only the question and the terms contained therein, but also the sources collected by the given model to answer the question play a significant role. In this context, it can be recommended that it is advisable to use prompts that are detailed and include restrictions or even exclusions. In the case of sources, I consider it a further necessary step to check the timeliness of the sources, as this has an influence on the answer.

With regard to the accuracy of references to legal provisions, *progress has been made* and significant improvements have been observed, but in many cases the response did not contain essential information, or the reference to the legal provision was not properly indicated in quotation marks, which could be significantly *misleading*. In this case, the operating mechanism by which certain generative AI systems do not display all text elements in the response is also unknown.

Compared to the preliminary examination, it was found that the generative AI systems examined are improving, but it was also highlighted that *certain types of errors occur in the same form* as in the preliminary examination, or in some cases in a different form. Nevertheless, in summary, there is still room for improvement in terms of consistency, source use, accuracy and the same types of errors: some terminological errors no longer appeared, but the same type of error occurred in relation to other terminology. In connection with tracking legislative changes, it should be noted that selecting the appropriate time frame for answering the question was difficult.

In addition to conscious and responsible use, taking into account the limitations of the generative AI systems examined, *they can also be used to support legal work*, provided the check is essential . The emphasis is not on which generative AI system is "the best", as the results have shown that there is no absolute winner among the generative AI systems examined. A single good performance does not mean that the next query will also yield a similarly positive result.

IX. Research and application limitations

The following research and application limitations apply to the research results presented in this dissertation:

- With regard to applications, the companies developing the AI-supported solutions featured in this dissertation are operational, and the AI-supported solutions they have developed are available at the time of writing, but it is possible that they may cease to exist or become unavailable over time.
- The availability and applicability of the models of the generative AI systems examined depend on the developing companies, so their availability and applicability are subject to change.
- The study was conducted on publicly available, general-purpose generative AI systems.
- Based on the case study, conclusions can be drawn at a given point in time with regard to the issues examined. Due to the nature of generative AI systems, different types of errors may occur when querying at different times.
- The case study referred exclusively to specific provisions of the Hungarian Labour Code.
- During the query on 19 May 2025, a new context window had to be opened for Claude 3.7 Sonnet due to the error message that appeared.

X. Results, recommendations

In this PhD dissertation, I examined the use of AI from a law-specific, practical approach. As a first step, the conceptual framework of the dissertation was defined, followed by an overview of the development of AI. As a result, it became apparent that in recent years, especially following the widespread adoption of generative AI systems, the use of AI has reached a new level. Following the rapid development of the technology, the AI Regulation came into force in the European Union as a result of a lengthy legislative process. The topicality of the subject is indicated by the fact that, despite the fact that certain provisions of the AI Regulation are not yet applicable, a proposal has already been made to simplify certain provisions. It is important to note that the AI Regulation also contains provisions on AI literacy, emphasising its role. At the same time*, it should be noted that there is currently no professionally accepted standard for AI literacy, and due to the development of AI systems, it requires continuous learning and self-improvement.*

In order to provide a comprehensive picture of how AI is used in legal work, I studied 12 international surveys and reports, which found that the use of AI in legal work is becoming increasingly important, with both law-specific and general (generative) AI systems being used. At the same time, the results of the surveys also pointed to gaps in the preparedness of legal professionals, with the development of AI skills emerging as a solution. In addition to presenting international trends, I also reviewed specific areas of application, for which I described specific AI-supported applications by category. With an emphasis on a practical approach, I reviewed various legal LLM benchmarks and the procedures for measuring their performance, which led me to conclude that it would be appropriate *to apply my own set of analytical criteria* to the issues I was examining.

Based on the results of my qualitative exploratory research, I concluded that the generative AI systems examined give the illusion of consistency, which is also a risk factor. In the course of my examination of specific legal provisions, In many cases, contradictions between responses within a single answer and between successive answers were found. In the context of resource utilisation, it is crucial which source a given generative AI system uses to respond, but in this case the "black box" effect prevails, as there is no information available on the selection mechanism or the ranking of sources. In the case of the generative AI systems examined that have a web search function, even if a suitable source appears among the sources, this does not mean that the generative AI system in question will use only or even the appropriate source for its response. It was also found that the use of official legal search interfaces was not necessarily preferred when responding.

Based on the results of the research, significant progress has been made in terms of the accuracy of references to legal provisions, but it is also apparent that in many cases, did not include essential missing elements, which could be misleading. The number of errors has decreased significantly compared to the previous situation, which may also mean that they are more difficult to notice and identify without the appropriate skills and expertise.

Due to the specific nature of the legislation, keeping track of legislative changes poses a challenge for the generative AI systems examined, but with proper oversight, these technological solutions can also be used in legal work in a conscious and responsible manner.

Considering all the responses to the research, there is no "best AI system" in terms of performance, as the results show that, despite significant progress, new types of errors are also appearing, and previously identified types of errors continue to occur in the same form. In my opinion, it is important to understand the limitations of these systems and take into account their operating mechanisms, but at the same time, they can play a supportive role in legal work.

During my research, I identified frequently occurring error types related to a specific topic, and then created my own set of analytical criteria as part of the methodology used. This set of criteria can also be applied to other text analysis studies.

Reflecting on the research questions, the following theses can be made:

(1) Although the various generative MI systems examined perform differently on individual questions, there is no generative MI system with the "best" overall performance.

(2) In the context of the generative AI systems examined, consistent responses remain an illusion and a source of risk. The generative AI systems examined are capable of producing accurate results even from inadequately used information or inadequate reasoning, and they are also capable of producing contradictory or incorrect results even from adequate sources. Different queries may yield answers of varying accuracy for the same question.

(3) The "black box" effect is characteristic of source management. In the case of the generative AI systems examined that have a web search function, even if an appropriate source appears among the sources, this does not mean that the generative AI system will (only) use the appropriate source to respond.

(4) Generative AI systems are also evolving, providing better and more accurate answers, but they return to the same pattern: certain types of errors can occur at any time.

(5) For the conscious and responsible use of generative AI systems, it is essential to understand the general operating mechanisms, limitations and "potential for error" of these systems, to continuously develop AI skills, and to exercise control.

Referring to the title of the dissertation, based on the results of the research, *consistency* in generative AI systems *remains an illusion.*

The analytical criteria, identified error types and findings can serve as guidelines for the scientific community – researchers and educators – as, according to the available information, no similar law-specific comparative text analysis and thus no examination of different generative AI systems from this perspective has been carried out in connection with Hungarian legislation. The analytical methodology described can facilitate the development of AI skills among both practising lawyers and law students.

The development of AI skills is of paramount importance both today and in the future. Today, it is still a competitive advantage, but in the future, the lack of these skills may become a disadvantage in the work of lawyers. In my opinion, the development of AI skills in legal education is justified and necessary. In view of all this, I propose that the responsible

use of AI be incorporated into legal education in a practice-oriented manner, if possible with the presentation of specific law-specific AI-based applications. This is important in view of the fact that today's law students will become the lawyers of tomorrow. It can be an active part of preparing for the changing and transforming labour market situation. Critical thinking is closely related to AI skills.

A further direction for research could be to repeat the questionnaire used as the basis for the empirical research in the future, thereby enabling the development of AI system models currently available at the time of future research to be examined. The research could also be expanded to include other legal provisions and legislative amendments.

XI. Publications on the topic of the PhD dissertation

- Kardos, Vivien: Data Protection Challenges in the Era of Artificial Intelligence, In: Thomas, Hemker; Robert, Müller-Török; Alexander, Prosser; Péter, Sasvári; Dona, Scola; Nicolae, Urs (szerk.) Central and Eastern European e|Dem and e|Gov Days 2021, Wien, Ausztria: Österreichische Computer Gesellschaft (ÖCG), 2022 (March). pp. 285-294. DOI: https://doi.org/10.24989/ocg.v341.21

- Kardos, Vivien: Privacy literacy and the protection of personal data in the mind of law students, PRO PUBLICO BONO: MAGYAR KÖZIGAZGATÁS; A NEMZETI KÖZSZOLGÁLATI EGYETEM KÖZIGAZGATÁS-TUDOMÁNYI SZAKMAI FOLYÓIRATA 9:4, 2021. pp. 124-141. DOI: 10.32575/ppb.2021.4.8

- Kardos, Vivien: Insight into the Perception of Personal Data Among Law Students. Central and Eastern European EDem and EGov Days 338: pp. 125-134., 2020 (July). DOI: https://doi.org/10.24989/ocg.v.338.10

- Kardos, Vivien Kata: Joghallgatók adatvédelmi tudatossága, FORUM: PUBLICATIONES DISCIPULORUM IURISPRUDENTIAE 3, 2021. pp. 291-322.