UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

The Calculated Law

Exploring the limits of transformer models in legal reasoning

William Backgård

Juridiska institutionen
Juristprogrammet HT24
HRO800 Examensarbete, 30 högskolepoäng
Handledare: Leila Brännström

Abstract

This thesis delves into the capabilities and limitations of Large Language Models (LLMs) used in legal reasoning, examining how these models process legal information to solve complex legal problems. A techno-legal analysis was conducted, comparing the technologies of LLMs to jurisprudence theories, to identify challenges they face in legal problem-solving. Also, an empirical evaluation was performed, assessing whether LLMs could handle a basic course in the Swedish law program.

The findings illustrate that LLMS excel in understanding legal language and are able to apply legal methodologies even in complex scenarios previously thought to be beyond the capabilities of AI. This includes more complex legal methods like analogies, and ethical considerations. However, they struggle with extensive problems requiring a step-by-step approach, where the problem needs to be broken down into smaller parts and addressed sequentially. Despite these limitations, the empirical test suggests that current LLMs could potentially pass a challenging law course. This raises concerns about academic integrity, as it becomes increasingly difficult to distinguish between student-written and AI-generated exam responses. The growing capability of LLMs in legal reasoning also calls into question their future role in the legal profession and their potential impact on the work of lawyers and judges.

To further elaborate, the ability of LLMs to potentially pass law courses may create difficulties in upholding academic integrity, as it becomes challenging to distinguish between student-written and AI-generated exam responses. The increasing capability of LLMs in legal reasoning also raises questions about their future role in the legal profession and how they might impact the work of lawyers and judges. This implies a potential shift in the legal landscape where LLMs could play a more prominent role in legal research, analysis, and decision-making. However, it also necessitates careful consideration of ethical implications and the need for human oversight to ensure fairness, accountability, and responsible use of LLMs in the legal domain.

Key words: Pre-trained Transformer Models, Legal reasoning, Legal Judge Predictions, Large language models, Legal methodology, American Legal Realism, Judicial hunch theory, Techno-legal analysis.

Acknowledgements

I would like to express my sincere gratitude to Leila Brännström, who has been my supervisor throughout this thesis and provided me with invaluable feedback. I also extend my heartfelt thanks to Jakob Heidbrink, Marianne Rödvei Aagaard, and David Dryselius for taking the time to participate in my evaluation of AI chatbots.

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."[1]

[1] This quote is often attributed to Marie Curie, but there is no definitive evidence that she actually said or wrote it.

Table of contents

Abbreviations

Term	Definition
AGI	Artificial General Intelligence – A level of artificial intelligence that exhibits human-like cognitive abilities across a broad range of tasks. Unlike narrow AI systems designed for specific functions, AGI aims to understand, learn, and apply knowledge in a generalized manner, adapting to new challenges much like a human would.
AI	Artificial Intelligence.
API	Application Programming Interface – A set of rules that allows software programs to communicate with each other. It acts as a bridge, enabling one program to integrate with and interact with another, for instance, by accessing data from an external database without needing to know its internal workings.
Chatbot	A chatbot in modern AI is an interactive software application that uses natural language processing and machine learning to simulate human-like conversation. Examples include ChatGPT, Gemini, and Claude, which can understand user input, generate context-aware responses, and perform a range of tasks from answering questions to assisting with various services.
CoT	Chain-of-Thought – A technique in AI that involves breaking down a problem into a series of intermediate reasoning steps before arriving at a final answer. This process makes the model's decision-making more transparent and often leads to more accurate outcomes by systematically processing complex information.
Fine tuning	The process of taking a pre-trained model and further training it on a smaller, specialized dataset. It refines the model's performance for a specific task, allowing it to adapt to new, domain-specific information without being built from scratch.
LJP	Legal Judge Predictions – The task of predicting the outcome of a court ruling in a real or hypothetical case.
LLM	Large Language Models – AI models with billions of weights and parameters, trained on vast amounts of text data. They are primarily used for generating and understanding text.
Matrix	A matrix is a rectangular arrangement of numbers or symbols, organized in rows and columns. It can be viewed as a table with information structured in a systematic way. The rows and columns are often their separate vector.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

The rows and columns are often their separate vector.

$$\begin{array}{cccccc} & \boldsymbol{w_1} & \boldsymbol{w_2} & \boldsymbol{w_3} & \\ & \parallel & \parallel & \parallel & \\ \mathbf{A} = & \left[\begin{matrix} 1 \\ 4 \\ 7 \end{matrix}\right. & \begin{matrix} 2 \\ 5 \\ 8 \end{matrix} & \left.\begin{matrix} 3 \\ 6 \\ 9 \end{matrix}\right] & \begin{matrix} =\boldsymbol{v_1} \\ =\boldsymbol{v_2} \\ =\boldsymbol{v_3} \end{matrix} \end{array}$$

$$\boldsymbol{v_1} = [1,2,3], \qquad \boldsymbol{v_2} = [4,5,6], \qquad \boldsymbol{v_3} = [7,8,9]$$

$$\boldsymbol{w_1} = [1,4,7], \qquad \boldsymbol{w_2} = [2,5,8], \qquad \boldsymbol{w_3} = [3,6,9]$$

A matris can also be of any length in both columns and rows:

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{a_{1,1}} & \cdots & \boldsymbol{a_{1,m}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{a_{n,1}} & \cdots & \boldsymbol{a_{n,m}} \end{bmatrix}$$

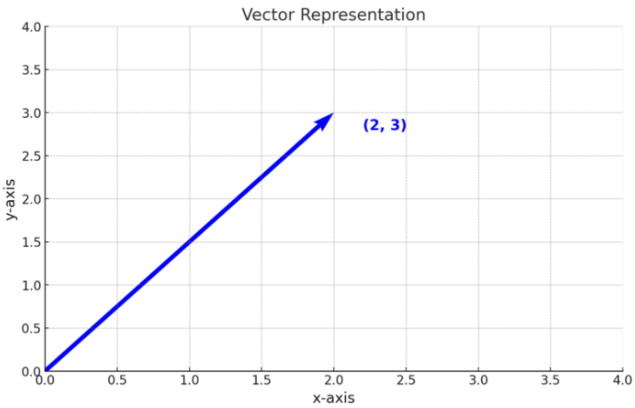
ML Machine Learning – A branch of artificial intelligence that uses data and algorithms to enable computers to learn and improve over time. It allows programs to recognize patterns and make decisions without being explicitly programmed for every task.

Modern AI Systems AI systems developed and released within the last five years. Most notably OpenAI's "ChatGPT", Google's "Gemini", Antprohic's "Calude". This is not, however, limited to chatbots.

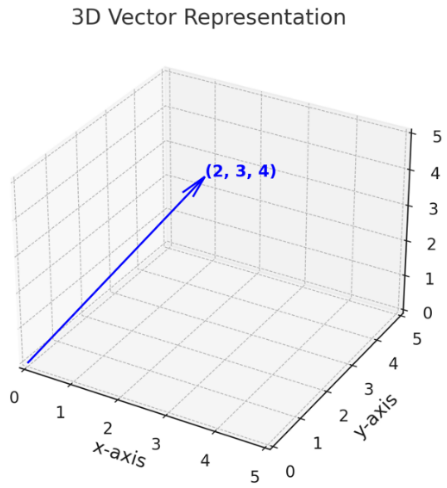
RAG Retrieval Augmented Generation – A technique that allows a language model to combine its own generation abilities with information retrieved from external sources. This means that the system can look up relevant data and then use that information to produce more accurate and detailed responses.

Parameter Often synonymous with weights, parameters refer to any numerical values used within an AI system.

Vector



A vector is a mathematical expression showing an arrow pointing to a certain point, in this example to the point (x = 2, y = 3) also notated as [2, 3]. This is shown on a 2D-plane. A vector may also be presented in a 1D-plane (everything being on the same line) then only having one coordinate (for example (x = 2)). A Vector may even be situated in a 3D-space as seen below.



The vector may even have a coordinate in the 4th dimension. While this is very difficult to imagine, it is much simpler in mathematical connotation (for example a 4th dimensional vector may be written as $[2, 3, 4, 6]$) . This can be done with any number of dimensions n. A vector in the n-dimension may be written as:

$$\mathbf{v} = [v_1, v_2, v_3, \ldots, v_n]$$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_3 \\ \cdots \\ v_n \end{bmatrix}$$

Weights In neural networks, this is typically part of the equation that occurs within the network. In its simplest form, it simply means that numbers are added with this weight.

XAI Explainable AI – AI systems and methodologies that provide transparent and interpretable results. XAI aims to demystify the decision-making process of AI models, enabling users to understand the rationale behind outputs and build trust in the technology.

1. Introduction

Throughout history, technological advancements have transformed the labor market. The Spinning Jenny drastically reduced the need for hand weavers, typewriters replaced scribes and copyists. These workers evolved into secretaries, whose roles diminished with the advent of modern PCs. Telephone exchanges rendered telephone operators obsolete, and digital cameras led to the decline of jobs in film development.

In the architectural field, the emergence of CAD[2] and other 3D software has shifted the focus from hand-drawn sketches to digital visualizations. Similarly, in banking, customer-facing roles have gained prominence over back-office administrative work.

We are now at the cusp of another significant technological leap. In the World Economic Forum, this has been called the "*Fourth Industrial Revolution*".[3] The rise of this rapid change may in large part be due to one invention: *transformers*.[4]

This brings us to a crucial observation: transformers may well be the Spinning Jenny of our era. They could lead to the obsolescence of certain occupations, as happened with telephone operators, or they might drive the evolution of roles as seen with architects. We find ourselves on an upward slope of technological progress, uncertain of our exact position or what lies ahead.

For decades, lawyers remained largely unaffected by technological evolution. The immense bank of written legal literature was not only a source of knowledge, but also a way of keeping knowledge out.[5] While PCs improved administrative tasks, the core of legal work remained unchanged. Technology primarily generates new clients rather than disrupting legal practice. For example, the rise of e-commerce created new challenges that had to be addressed legally. Although the internet made information more accessible, legal sources remained complex for outsiders and often exclusive to lawyers.

Law cannot be simply automated. It deals fundamentally with human life and relationships. Understanding law requires not just access to legal sources, but comprehension of how people think, feel, and interact. Law is inherently linguistic — to master law, one must master language. This has long been the prevailing belief.

---

2 "Computer-aided design". This is normally a tool used by architects and engineer for visualization of structures, tools etc. These tools often can calculate physics.

3 "The Fourth Industrial Revolution: what it means, how to respond," World Economiv Forum, 2016, accessed 15 November 2024, https://www.weforum.org/stories/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/.

4 "Transformers Revolutionized AI. What Will Replace Them?," Forbes, 2023, accessed 25 november 2024, https://www.forbes.com/sites/robtoews/2023/09/03/transformers-revolutionized-ai-what-will-replace-them/.

5 See Kronblad who argues that there are several reasons why lawyers resist digitalization, including their desire to protect their knowledge capital. Kronblad, Pregmark and Berggren, *Difficulties to digitalize: ambidexterity challenges in law firms*.

The emergence of pre-trained transformer models challenges this paradigm. For the first time since AI's inception in the 1950s, we have systems that genuinely understand language, with some arguing they exceed the average human's linguistic capability.[6]

2. Problem and purpose

Transformers enable AI systems to excel in understanding language. Modern chatbots are often described as generative, meaning they have one primary task: to predict the next word of a text. To accomplish this, these systems use multiple layers of transformers and embedding layers that employ millions of parameters to calculate the most probable next word. This is essentially all that modern chatbots do.[7]

The work of legal professionals is broad and encompasses many different forms of practice. The essence of what law is, should be reflected in the skills that students acquire through law school, as well as in the primary professional activities of those who are called "lawyers". This essence will be known as *the art of law*.

At first glance, one might assume that these transformer models could not possibly master the *art of law*. After all, lawyers need to do much more than just write one probable word at a time. However, as will become evident, these AI models can perform many tasks previously believed to be exclusive to human lawyers. While it is true that AI only predicts the next most likely word, this description is somewhat misleading. There are numerous complex functions behind the layers that make this seemingly magical process happen.[8]

Due to the emergence of tools that could challenge the work of traditional legal professionals, it is therefore worthwhile to examine this further. There is also strong interest from academic circles, as these tools may completely transform legal education. A field that required legal methodology was considered almost impossible to cheat through. With the advent of LLMs, this could be completely overturned. Therefore, this thesis examines the capabilities of transformer models in the field of law. The focus is on understanding how these models process legal reasoning and handle complex legal problems.

The focus will be on the methodology within legal practice. Similar to how a computer program transforms input into output through its functions, a lawyer processes legal problems to produce solutions through legal methodology. This thesis will investigate whether this legal methodology can be represented as a function, or as a complex series of functions in the form of a LLM.

While exploring these questions, I will inevitably touch upon the field of explainable AI (XAI), the science of understanding how AI arrives at its decisions. Although XAI is not the primary

[6] Bojic, Kovacevic and Cabarkapa, *GPT-4 Surpassing Human Performance in Linguistic Pragmatics*.
[7] See 6. The computations in large language models
[8] This will be explained in 6. The computations in large language models and more specifically The hunch theory in AI-models

focus of this research, the nature of legal reasoning itself, which relies heavily on clear argumentation and persuasion through language, means that any AI system demonstrating mastery of legal reasoning would inherently need to explain its thought process effectively.

The final questions to be answered in this thesis are as follows:

1. *Are there any challenges within Legal Problem Solving that LLM-based AI models cannot or find very difficult to solve?*

This is primarily a theoretical question, where I want to investigate the limits of AI capabilities within the field of law. This study has an extremely broad scope, primarily focusing on the jurisprudence taught during law school education. However, it addresses more advanced reasoning beyond simply applying clear rules. Specifically, it examines how to handle unregulated areas, ethically challenging problems, complex trade-offs, and the fundamental challenge of identifying what constitutes a legal problem in the first place, including the methodology and sequential steps required to resolve it.

To anchor theoretical investigations in reality, I have also chosen to develop a second question that, as objectively as possible, should be able to determine where LLMs' capabilities within law actually stand today, namely:

2. Are LLM-based AI models capable of handling a basic course in the Swedish law program?

More specifically, the question is whether it can handle HRO201 - Obligationsrätt (Contract Law). The course deals with a considerable amount of legal methodology and is therefore not just a matter of answering pure factual questions.

3. Method

To answer my research questions, I have chosen to divide the methodology into four distinct methods.

The first two methods are primarily *descriptive* and serve to establish an underlying theoretical foundation for the subject matter. These methods address the theoretical foundations within jurisprudence (3.1), and the field of computer science, particularly regarding Large Language Models (3.2). This framework serves as the foundation for addressing questions 1 and 2. Since my research explores legal methodology and law on a meta-level, it is essential to anchor the analysis in legal theory to assess LLMs through a legal lens. The descriptive method for LLMs is primarily included to give both myself and the reader a necessary understanding of the technology, serving as a foundation for the methodological approach introduced in the next paragraph.

To answer question 1, a "*techno-legal analysis*" will be conducted where, the technologies of LLM are compared to the theories of jurisprudence to answer the specific question (3.3). This method is used to evaluate the capabilities of LLMs within the field of law.

Finally, to address the last question, a mostly *objective empirical method* will be used, much like traditional science articles (3.4). Here, I will present a research evaluation examining whether LLMs can objectively emulate legal professionals.

3.1. Theoretical foundations of LJP

Legal methodology is complex and encompasses several conflicting theories. To evaluate LLMs in this context, a clear theoretical framework is necessary to define legal methodology itself. Without such a framework, it is impossible to determine whether LLMs align with legal reasoning. This makes the assessment especially challenging when applying legal theory. For example, using natural law as a foundation would require comparing the functionality and output of LLMs to morally "correct" answers.[9] But whose "correct" answers should be used, and how they should be evaluated become an issue. The purely formalistic branch of legal positivism may lead us a step further, but also presents challenges. It builds on the idea of an internal structure of legal sources, but it would still be subjective and difficult to evaluate whether an LLM has reached the "correct conclusion."[10] All these legal theories contain normative elements, raising questions about which normative values to consider and why.

A more practical approach (and likely the only one of the possible) might be to compare the functionality of LLMs with a descriptive theory like American Legal Realism, which defines law and legal methodology by focusing on their actual practical effects and applications.[11] This provides concrete elements for comparison. This approach eliminates the need for subjective theories about whether the LLM's conclusions are "correct." Instead, we only need to determine whether the outputs align with those of practicing lawyers and judges.

With this theory as a foundation, I will be able to systematically develop and analyze LLMs and their outputs in relation to what constitutes law in each specific case. This framework will be particularly useful when comparing LLM performance to how legal professionals apply the law, as it provides a concrete descriptive legal reference. Additionally, it will be instrumental in addressing my later research question on how well LLMs perform in a law school course.

The primary sources for this methodological approach will be American Legal Realists. However, since American Legal Realism was developed within a fundamentally different legal system, I will supplement it with other related theories that align with its core principles. This

[9] Cf. how Strömholm et al. jointly define natural law according to Strömholm, *Rätt, rättskällor och rättstillämpning : en lärobok i allmän rättslära*, 88.
[10] Kelsen and Jestaedt, Reine Rechtslehre : Einleitung in die rechtswissenschaftliche Problematik (Studienausgabe der 1. Auflage 1934), 74–76.
[11] See 5.4. American Legal Realism.

will allow me to establish a theoretical foundation that is not exclusively tied to a common law system.

A broader complementary theory I will incorporate is polyvalence. This framework will help demonstrate that LLMs can adapt effectively to legal systems that are more complex and interconnected with other fields. This adaptability directly addresses a common criticism of AI in legal context, namely, that AI struggles to handle the intricate and multidisciplinary nature of legal reasoning.

The theoretical foundation used to address the research question 1 and 2 will therefore be based on American Legal Realism and partly on polyvalence. This allows this thesis to more objectively determine the correct solution to a legal problem and explain why LLMs are more suited for the legal world than many may believe.

3.2. Theoretical foundations of LLMs

LLM is a very new technology with limited exposure to legal circles. Therefore, this will require an interdisciplinary approach. It means I will create a theoretical framework for how these LLMs function in theory. This will involve highly technical and mathematical theoretical concepts.

Historically, many legal fields began as interdisciplinary studies before becoming established disciplines within law. For instance, environmental law was once a narrow field but evolved to encompass broader areas such as international environmental issues.[12] Similarly, we are likely at the dawn of "*AI law*." This field currently exists largely outside traditional legal boundaries, with the first AI directive only being enacted in 2024.[13]

Therefore, this theory needs to be unusually mathematical and technical compared to other theories found in legal articles. However, this will prove beneficial when examining and comparing LLMs with legal theory to answer question 1.

A significant challenge in this field is finding reliable sources. The availability of sources is limited for several reasons. First, LLM technology is relatively new, resulting in limited academic literature on the subject. Second, the field is highly commercialized, with leading researchers working for major companies like Google, Meta, and OpenAI rather than universities. As a natural consequence, the dissemination of knowledge is not as extensive as it would be if universities were the primary actors. Finally, much of the available literature is simply too complex to address within the scope of this thesis. For these reasons, I will draw from a broad range of sources, some quite complex, as the foundation for this theoretical framework.

---

[12] Michanek, *Den svenska miljörätten*, 60–81.
[13] ”AI Act enters into force,” 2024, accessed 29 november, 2024, https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en.

3.3. Theoretical challenges in AI-based legal reasoning

To address question 1, the theoretical framework from legal studies will be applied to the theoretical description of LLMs. Once there is a clear understanding of both legal methodology in this thesis and how LLMs function, these will be compared to determine how AI models actually solve legal problems in comparison to lawyers. This analysis will then progress to increasingly complex cases until the limitations of AI models become apparent. In other words, I will examine whether there are any aspects of a lawyer's work within Legal Problem Solving (LJP) that become difficult or impossible for LLMs, based on our understanding of how LLMs function. This is what is described as the *techno-legal analysis*.

Unlike question 2, this part will explore LLMs' capabilities and limitations on a more theoretical level. While I will employ actual LLM-based chatbots to illustrate both successes and failures in LJP, I will not restrict my analysis solely to their current performance. Instead, I will evaluate, using theoretical frameworks, whether handling various aspects of LJP is straightforward, challenging, theoretically possible, or impossible for LLMs.

I will examine common challenges within LJP, such as solving problems through analogies, proportionality assessments, and handling ethical considerations. A significant focus will be on how well AI handles complexity — particularly legal cases involving extensive information rather than isolated problems that humans might find challenging.

While AI models are evolving so rapidly that this thesis might become outdated shortly after being written, this methodology for question 1 can still provide valuable insights into how LLMs handle LJP, even if current LLM models are not yet capable of all these functions.

Once I have applied this method, I will have a clear understanding of which legal problems LLMs can theoretically handle, and which challenges they are likely to continue struggling with. The scope of this method is limited, as described in 3.4 Limitations.

3.4. LLM and legal education: an empirical evaluation

In research question 2 I aim to determine whether LLMs have the capability to pass a course in the Swedish law program. My methodology involves creating an empirical approach where I examine objectively how many AI-generated answers have the potential to achieve a passing grade in a course.

The AI models being tested will primarily be OpenAI's ChatGPT 4 and Claude's 3.5 Sonnet. These chatbots will be given access to the same materials that students have access to during an exam, as well as an exam question to answer.

One challenge is determining which of the AI-generated answers should be considered passing. If I were to personally evaluate what constitutes a passing answer, this would be highly subjective. Therefore, I instead apply a more objective empirical method that was first conceived by Alan Turing.

Alan Turing, often described as the father of artificial intelligence, was particularly interested in machines and their capacity for human-like thinking. He developed the now-famous *Turing Test* to evaluate whether machines can think. The test determines if a human can distinguish between interactions with a human versus a machine. If humans cannot distinguish between the two, then the machine is considered equivalent in terms of thinking capability.[14]

This method can be effectively applied to my problem. To evaluate whether an AI system's output matches that of a lawyer or judge, we can construct a corresponding Turing test. If an observer cannot distinguish between outputs from a lawyer and those from an AI system, we can conclude that the AI can reason like a lawyer.[15]

It is worth noting that this method of determining AI's cognitive abilities has faced some criticism, notably by the *Chinese room* argument.[16] However, this critique has itself been subject to substantial counterarguments.[17] The Turing test provides this thesis with its final methodology: an *empirical approach*. Through the combination of American Legal Realism theory and the Turing test, we can objectively determine whether an AI system can function like a lawyer and identify any limitations in its capabilities.

This method of comparing legal applications between an AI system and a judge is, however, very complicated in practice. The AI system would need access to all materials available to a judge, and even if this were practically possible to obtain, it would be too much to input into a current chatbot. Furthermore, this would require feeding a lot of images and audio into the AI system. Although progress is being made, no current chatbot can handle these different types of media as effectively as needed. The most effective ones only handle text.

Moreover, a judge's output, the judgment, is considerably longer than what a modern chatbot can create. This means that currently, we must be content with handling one part of the problem at a time, for example, one relationship at a time.

---

[14] ”The Enigma of Alan Turing,” 2015, accessed 10 october, 2024, https://www.cia.gov/stories/story/the-enigma-of-alan-turing/. See also Turing, *Computing Machinery and Intelligence*. Where Turing firstly describes his connection with "thinking machines", philosophy and psychology. Ibid., ibid.

[15] According to Turing. There is also a philosophical explanation for why this must be the case. If an external agent has no means of distinguishing one thing from another, they are, in a philosophical sense, considered identical. Similarly, exams in legal education can be seen as effective tests designed to differentiate students who are well-prepared for legal practice from those who are not.

[16] John Searle argued that computers cannot truly understand or possess consciousness, even if they appear intelligent. In his Chinese Room thought experiment, he imagines a person inside a room following a set of rules to manipulate Chinese symbols without actually understanding the language. From the outside, it seems as though the person understands Chinese, but they are merely processing symbols based on instructions. Searle uses this to claim that computers, like the person in the room, operate on syntax rather than semantics, and therefore lack true understanding regardless of how convincingly they simulate intelligence. See Häggström, *Tänkande maskiner*, 229.

[17] Understanding language involves more than manipulating its structure (syntax); it also requires grasping its meaning (semantics). Semantics entails connecting words and expressions to concepts, ideas, and contexts, which many would argue is a fundamental aspect of intelligence. If an AI can use and interpret language in a way that demonstrates a deep understanding of its semantic relationships, it could be argued that this constitutes a form of intelligence, even if it differs from human consciousness. See Mitchell, *Artificial intelligence : a guide for thinking humans*, 49.

Finally, it is quite easy to conclude that current AI systems cannot come close to producing the judgments that real judges create. This is partly because these judgments are based on sources that AI systems do not have access to, and partly because the logic and application in these judgments are at a much higher level than modern AI systems can achieve today.

Instead, this thesis will conduct an empirical attempt to see if AI robots are heading in that direction. I have accessed the exams and exam answers from a specific course during the second semester of the law program in Gothenburg. The course is called "*HRO201 – Obligationsrätt*" and covers the fundamental jurisprudence within civil law (at other institutions, this is called Civilrätt 1, "Civil Law 1").

There are several reasons why I have chosen this particular course. In my initial tests, it appears to have a reasonable level of complexity. Secondly, the course incorporates several advanced legal elements, such as analogical reasoning and contract interpretation, unlike more basic courses that primarily focus on the direct application of statutory text.[18] Thirdly, this course is often regarded as one of the most challenging for law students. As a result, the stakes are higher, making the incentive for students to use AI technology to "cheat" more significant.

I will therefore conduct a Turing test on an exam in this course. This will determine whether AI can emulate students who achieve the highest grades in this course. This is particularly relevant in our current times, as there have been several instances of students cheating on exams using chatbots.[19] If the results are indistinguishable, it would indicate that AI is approaching the same level we expect from law students after their first year in the law program.

It is worth noting that not all students pass the course. This means that even if an AI model can pass as a student, meaning that it is impossible to distinguish between AI-written and student-written answers, the AI model might still produce answers that would not meet the passing criteria. To address this issue, the lecturers will evaluate the quality of the answers using the same grading system as used in the course.

However, grading answers requires significantly more time than I can reasonably ask of the lecturers. Additionally, more than half of the lecturers are not teaching this specific course (although they work in the same legal field) and therefore do not have detailed insight into the course's grading criteria. As a result, while this test provides a good indication of the quality of

[18] The precise arguments to be used in each area are not strictly defined within Swedish legal doctrine. Nääv and Zamboni, *Juridisk metodlära*, 36–37. Building an understanding without clear rules for which arguments to apply in specific situations presents an equal challenge for both AI systems and law students.

[19] See "Göteborgs universitet läxar upp student – stängs av efter misstänkt AI-fusk," Göteborgs-Posten, 2024, 20 November, 2024, https://www.gp.se/nyheter/goteborg/goteborgs-universitet-laxar-upp-student-stangs-av-efter-misstankt-ai-fusk-.32933d07-be22-4e38-8b08-ed95f8a95272., "Student använde AI på tenta – frias från anklagelser om fusk," Göteborgs-Posten, 2024, accessed 21 November, 2024, https://www.gp.se/nyheter/goteborg/student-anvande-ai-pa-tenta-frias-fran-anklagelser-om-fusk.47b3b9bb-96a2-4b6e-8131-86fd2be1fa50. This is particularly noteworthy as students have found ways to bypass the "Safe Exam Browser" protection during in-person exams. This means students can access content outside the exam environment and potentially use chatbots even during supervised examinations.

AI-written answers, it cannot definitively determine whether AI models would pass the course. Such a conclusion would require a more comprehensive test.

3.5. Limitations

A lawyer's work encompasses many social aspects, such as communicating with clients, other lawyers, and so forth. While this is an important part of legal work, it will not be researched in this thesis. Furthermore, the theoretical work contains many different aspects. To narrow the scope of this thesis, the focus will be on LJP. This involves attempting to predict how a judge will rule in a case. This is the primary skill practiced by students in law programs. Strong LJP ability is also fundamental to all other legal work. Even if legal work does not directly involve predicting how a judge will rule, lawyers must always keep in mind that their work may end up in court.

The subject of LLM in law can obviously be vast, as combining modern AI systems with law encompasses multiple domains. However, to best address this question, it is important to maintain a broad perspective. This is an area that has not been extensively researched yet, which is why this investigation needs to start broad before becoming more focused.

Since my purpose is to find the boundaries of what AI cannot do, my limitation will be in trying to identify these constraints. The limitation thus becomes finding the line where the difficulty level becomes too high for AI rather than which different areas of law it can handle.

Despite this broad view, I have chosen to limit my legal investigations primarily to civil law, specifically civil law where relationships are mostly limited to two parties. Priority of claims issues[20] therefore do not form part of the problem scope. Civil law remains a complex legal system that includes direct applications, analogies, and teleological applications, while also containing numerous ethical aspects and having strong connections to what one might classify as non-directly legal, that is, reality.

A significant limitation concerns bias and discrimination in AI systems. This is an area that typically receives substantial attention from legal scholars examining AI systems. However, investigating bias in AI systems presents numerous complex challenges. Therefore, while this area is extremely important and requires continued research, it falls outside the scope of this thesis.[21]

Evidence assessment constitutes a significant aspect of law that receives minimal attention during law school education. Given its limited coverage in the law program, this thesis will not

[20] Often called “sakrätt” in Swedish law.

[21] Addressing bias in AI systems is a distinct topic that can be divided into several key questions. First, there is the normative question of how one should act when using AI in legal contexts. For example, how to handle affirmative action. There are also philosophical questions about the nature of discrimination and how it differs from direct decision-making. Finally, there are pure logical and mathematical challenges in this field. For instance, mathematical proofs demonstrate that it is impossible to create AI systems that are completely free from any form of bias. See Kleinberg, Mullainathan and Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*.

focus on these aspects. However, similar methodologies from this thesis could be applied to address questions regarding evidence handling and evaluation.

The course that the Turing test was applied to is a so-called portfolio course, which means that students create a portfolio at the end of the course where they demonstrate that they have achieved all the learning objectives throughout the course. However, the greatest challenge of the course is writing the exams, with the final exam covering the entire course content. Additionally, other institutions do not use this portfolio approach, but only exams. Therefore, I will treat the tested course as if it consists only of one exam test.

4. Previous research

Most of the research in the field of "AI and law", and more specifically "LLM and law", originates from computer science, which is natural since AI was developed within computer science. Research from legal science perspectives is more limited. In my interpretation, this is due to two factors. First, legal science encompasses much more than just AI. Understanding modern AI systems requires a deep understanding of the underlying technology; a legal researcher cannot analyze a system without understanding how it works.[22] Additionally, I believe there is likely a prevalent skepticism towards AI within the legal community. From my perspective, law is fundamentally a field created by humans for humans, and this may explain the reservation many legal professionals have towards AI.

There is one area within AI that has been extensively researched from a legal perspective, namely AI systems' tendency towards bias. This can be seen as part of the skepticism towards AI in the legal field. It is an area where it is particularly easy to criticize the use of AI in legal contexts. However, I want to clarify that while bias in AI systems is a real concern, this thesis will not focus on this issue.

Previous research must be divided into *pre-transformer* and *post-transformer* periods, as they are distinctly different. Most pre-transformer research indicates that researchers did not anticipate how quickly modern AI's potential would emerge, and many of the problems they described with AI systems are either no longer relevant or not as fundamental as their research suggested.

4.1. Pre-transformers legal AI research from the field of computer science

In 1990, computer science doctoral student Rissland investigated AI in relation to law. She developed a theoretical model for handling legal matters and identified numerous challenges in creating such a system. Rissland demonstrated that the legal system is highly complex when

[22] Much of the previous AI research focuses on robotics, that is, rules for robots, despite the fact that the AI most relevant to legal work does not contain any robotic elements or mechanical parts that physically interact with the world. See example Pasquale, *New Laws of Robotics*. Turner, *Robot rules : regulating artificial intelligence*. Abbott, *The reasonable robot : artificial intelligence and the law*.

it comes to handling legal problems. Furthermore, such a system would need to be able to analyze and apply precedent cases. Additionally, AI in law becomes problematic when dealing with vague rules.[23]

In 1997, computer scientist Brench-Capon proposed a theoretical approach to constructing an AI system capable of handling legal argumentation. This approach is based on a purely rule-based legal framework with a clear logical structure, which he argues could be modeled by an AI system.[24]

4.2. Pre-transformers legal AI research from the field of jurisprudence

Famous professor of legal informatics Sartor, and professor of artificial intelligence and Law Prakken, maintains a positive outlook on AI's potential as a testing environment for legal rules and reasoning. However, they emphasize that law cannot be reduced to simple logical rules, as it often comprises conflicting arguments and underlying values that must be weighed against each other. A purely logical legal system risks creating contradictions and unexpected outcomes, particularly since the application of laws often requires interpretation of their purpose and the societal values they are intended to protect. They argue that the development of AI systems in law must account for this complexity and value-based nature, while integrating models that handle dynamics, exceptions, and procedural aspects in legal processes.[25]

In a previous article, Sartor reiterates that law is value-based and goal-oriented. He demonstrates that law involves many trade-offs between these aspects. Interestingly, he presents several mathematical and logical models as tools to analyze and structure these trade-offs.[26]

Professor Scherer expresses skepticism towards AI's ability to perform legal work, emphasizing that current AI systems face fundamental limitations in legal reasoning. He points out that AI's reliance on historical data and probabilistic models makes it ill-equipped for novel legal situations. Moreover, Scherer highlights AI's critical shortcomings in providing transparent legal reasoning, understanding human factors, and grasping complex principles of social justice, all essential aspects of legal practice.[27]

There are also research findings from Aletras, Tsarapatsanis, Preoţiuc-Pietro and Lampos from 2016 regarding the prediction of outcomes from the European Court of Human Rights (ECtHR). They created an algorithm that achieved up to 79% accuracy. However, it could only produce a binary answer (violation or no violation). Additionally, it only addressed centered

[23] Rissland, *Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning*, 1957–1973.
[24] Bench-Capon, *Argument in Artificial Intelligence and Law*.
[25] Prakken and Sartor, *Law and logic: A review from an argumentation perspective*, 215–217, 241.
[26] Sartor, *Doing justice to rights and values: teleological reasoning and proportionality*, 177 pp.
[27] Scherer, *Artificial Intelligence and Legal Decision-Making: The Wide Open?*, 542–543, 555.

questions, i.e., whether specific circumstances violated a particular right. In this way, this is much simpler than what will be investigated in this thesis.[28]

Wahlgren, while not stating this in a research paper, also concludes that automated decision-making requires a formalization of language, i.e., breaking it down into if-then statements, in order to function. Wahlgren describes it as the requirement to translate decision-making materials into a formal representation to enable automated decisions.[29]

4.3. Post-transformers legal AI research from the field of computer science

Computer scientist Kang et al. have studied AI applications in law. This research is most closely related to my own research questions. They found that AI can solve simple legal problems as long as they are sufficiently well-defined. They also discovered that AI systems struggled with larger problems requiring systematic techniques to solve. However, they noted that this limitation can be addressed by breaking down the problem into smaller components before having AI address each part. However, the research is limited, as it focuses only on Malaysian contract law and Australian social law, which may differ significantly from Scandinavian regulations and are relatively rudimentary in comparison. The research is also from 2023, and the chatbots have evolved a lot since then. The questions seem to only be testing the chatbot in the simplest judicial problems (no unclear regulations etc.).[30]

4.4. Post-transformers legal AI research from the field of jurisprudence

Sartor, together with Rotolo, returned to the field to focus on legal argumentation, particularly the distinction between explaining and justifying decisions. They demonstrate that this distinction is pragmatic rather than structural, meaning that while these two processes differ depending on the audience and purpose, they fundamentally share the same logical foundation. They argue that this perspective makes explainable AI (XAI) particularly crucial in the legal domain, as it is necessary both for making AI decisions comprehensible and for justifying them according to legal standards. This leads to the conclusion that AI systems capable of explaining their decisions are one step closer to being able to apply and justify legal principles.[31]

4.5. Conclusion of previous research

The review of previous research reveals several key findings. There is a dichotomy in perspective regarding AI in law. Computer scientists and researchers from related fields generally maintain an optimistic outlook on AI's future applications in law. Legal scholars,

---

[28] Aletras, Tsarapatsanis, Preotiuc-Pietro and Lampos, *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*.
[29] Nääv and Zamboni, *Juridisk metodlära*, 412—413.
[30] Kang, Qu, Lay-Ki, Trakic, Terry Yue, Emerton and Grant, *Can ChatGPT Perform Reasoning Using the IRAC Method in Analyzing Legal Scenarios Like a Lawyer?*
[31] Rotolo and Sartor, *Argumentation and explanation in the law*.

conversely, express reservations and highlight various challenges. This reflects a natural pattern where computer scientists emphasize the capabilities of their field, while legal scholars stress the complexity of their domain.

I argue that this opposition stems from misunderstandings on both sides. Generally, critics from the legal field fail to grasp the complexity that modern AI systems can handle, while computer science research tends to underestimate the intricacy of legal practice.

Profession	View of the law	View of AI
Lawyers	• A complex, nuanced system influenced by societal, cultural, and contextual factors. • Justice and objectivity are rarely absolute, with answers often dependent on circumstances.	• A logical machine designed to handle scenarios systematically through explicit functions. • Expected to provide predictable and explainable decisions.
Data professionals	• A closed, logical system where every rule has its place. • Outcomes can be derived deterministically, even if complex.	• A non-linear, intricate system lacking clear logic in how it reaches decisions. • Entangled with vast domains, building its reasoning statistically rather than rule based.[32]

Table 4.1 Lawyers and Data professionals corresponding view of law and AI.

During the pre-transformer era when AI was purely symbolic[33], skepticism towards AI's ability to understand and apply law was natural. After all, law is an extremely complex system with many aspects that are difficult to solve from a purely functional perspective. However, history shows that it is ill-advised to make definitive statements about what AI cannot achieve. The constantly emerging technologies in AI research are forcing legal professionals to reevaluate what should be considered "impossible for AI to achieve".

5. Theoretical framework

Many contrasting theories attempt to define "*the law*". A key distinction exists between theories that address what *law is* and those that consider what law *ought to be*. For the question at hand, understanding what law is becomes more relevant. This thesis will compare the legal work performed by a lawyer with that of an AI system.

[32] True for non-symbolic AI.
[33] Will be explained in 6. The computations in LLMs.

5.1. Legal positivism and formalism

Legal positivism originated in the 18th century Enlightenment, primarily through Jeremy Bentham's ideas. Bentham emphasized the importance of separating law from morality and argued that law's validity is based not on its ethical value but on its creation through legislation and its ability to govern human behavior. According to Bentham, law consists of rules reinforced by rewards and punishments, aiming to maximize social utility. This utilitarian view of law laid the foundation for a scientific analysis of jurisprudence, deliberately excluding moral and metaphysical elements.[34]

In modern times, these ideas were developed further by H.L.A. Hart, who refined legal positivism by introducing a theory of primary and secondary rules. Primary rules directly govern behavior, while secondary rules function as meta-rules for how the legal system is created, changed, and applied. A central concept in Hart's theory is "*the rule of recognition*," which defines what counts as valid law within a particular legal system. This rule provides the legal system with its legitimacy and connects it to societal institutions.[35]

A particular branch of legal positivism is *formalism*. It is important to understand this theory, as there can sometimes be a naive perception of law as a formalistic system. In this sense, formalism stands in contrast to more modern theories such as polyvalence. Hereafter, formalism should be understood as a monocentric system without contradictions. That is, a system where everything can be broken down into its components to later build a logical and completely correct conclusion.[36]

It is likely that no academics in jurisprudence would call themselves pure formalists as described above. However, it is unclear whether this is due to the fact that it is impossible to apply in reality. Our reality is far too complex to be described in a completely formalistic system.[37] Nevertheless, it does not rule out the theoretical possibility, but since it is entirely impractical in reality, it is not worth pursuing.

A common criticism of AI systems in law often stems from the assumption that the world must be reducible to pure formalism. An algorithm relies on if-then statements that require clear definitions without unambiguity, and therefore this formalism must exist. Upcoming theories

[34] Wacks, *Understanding jurisprudence : an introduction to legal theory*, 82–84. Strömholm, *Rätt, rättskällor och rättstillämpning : en lärobok i allmän rättslära*, 92.
[35] Wacks, *Understanding jurisprudence : an introduction to legal theory*, 95–102.
[36] This is partly based on theories from Kelsen's "Pure Theory of Law." The goal was to analyze law as an independent and coherent system, separate from morality, politics, and other social sciences. Kelsen viewed law as a hierarchical system of norms where each norm derives its validity from a higher norm, with a basic norm ("*Grundnorm*") serving as the ultimate foundation. Ibid 107. However, Kelsen himself acknowledges that in reality, there are cases that are indeterminate, and judges must make discretionary judgments. Nääv and Zamboni, *Juridisk metodlära*, 63–64.
[37] This can be compared to Dworkin's "Hercules," who can always arrive at the correct legal conclusion by sorting through laws and principles. Wacks, *Understanding jurisprudence : an introduction to legal theory*, 146–148. The analogy suggests that even if it is impossible in practice, there is always a correct answer. Similarly, one can imagine a supercomputer that can sort through all facts and regulations to arrive at the most obvious and non-contradictory solution on some level.

will differ from this formalism, and I will later demonstrate why LLMs do not require this formalistic worldview.

5.2. Legal realism

Legal realism is considered a form of positivism because it views law as separate from morality and as a social construction, but it differs in its perspective on the legal system itself. While pure positivists may see law as a closed system that can be examined in isolation, realists reject this view, arguing that law must be examined in its real-world context. Therefore, realists take an empirical approach, whereas positivists remain focused on abstract, formal principles and meta-rules that define legal systems.[38]

Legal realism can generally be divided into two subgroups: Scandinavian Realism and American Realism. Both of these are, as their names suggest, realistic in their view of what law is. The main difference lies in their methodology. While the Scandinavians emphasize social utility ("*samhällsnytta*") and maintain some philosophical methods, keeping some of the "ought to be" in their discourse, American realism is much more empirical in nature.[39]

5.3. American Legal Realism

The theory of legal realism takes, seemingly, a very pragmatic stance. The law is what the courts do, *nothing more pretentious*.[40]. Although one might believe that law is founded on axioms, logic, and ethical principles, it remains uncertain how these concepts are meaningful if courts do not reflect them in practice. Therefore, American Legal Realism is a pragmatic view and a descriptive model. It does not try to explain what the law ought to be.

This naturally opens the door for cases to be influenced and judged more in line with the judge's perspective. Their own moral and political views will color the outcome. It also shows how a judge can adapt a ruling based on social needs rather than strictly following legal or historical arguments.[41]

While it is true that American Legal Realism focuses heavily on the courts, it is more of a sociological theory, since it acknowledges the unimportance of theoretical terms within the legal system. A theoretical legal system that has no effect on the world is therefore not the law. The American Legal Realistic system is, after all, realistic.[42]

However, it would be a mistake to assume that American Legal Realism is solely concerned with courts. There are many other institutions in society that operate in similar ways, a fact acknowledged by American Legal Realists. Statesmen and legislators are also included in this framework and are, in a sense, judges within their own domains. These individuals have the

[38] Ibid 173–174.
[39] Ibid.
[40] Oliver Wendell Holmes, *The Path Of The Law*, 4.
[41] See Nelson demonstrating how this became evident in the American case Brown v. Board of Education Nelson, *Brown v. Board of Education and the Jurisprudense of Legal Realism*, 803–804.
[42] Wacks, *Understanding jurisprudence : an introduction to legal theory*, 174.

ability, albeit not as extensive as American judges, to influence people in ways similar to a court.[43]

5.3.1. Judicial Hunch Theroy

There are more aspects of American Legal Realism worth considering as a theoretical base for analyzing LLMs doing legal reasoning. Even though most of the profound American Legal Realists were active almost 100 years ago, a time before AI was even in the vocabulary, some of this would fit really well into the theory of combining AI with the art of law.[44] One of the most radical legal scholars and jurists was Jerome Frank. Frank studied the theory of facts, which falls within the category of epistemology. He questioned the objectivity of facts, arguing that their interpretation was largely dependent on judges and juries. The role of the court is to produce facts, but Frank saw this as absurd, pointing to the fact that a court case could be ruled completely different depending on which judge was to rule that exact day, and what time of the day it would be ruled.[45]

Frank argues that this occurs due to the subjective role of the judge in the ruling. The judge may unconsciously pick up on elements within the courtroom that influence their decision-making. This affects the outcome in one direction or another, leaving us uncertain as to whether truth can truly stand on such foundations. Frank argued this would be absurd. Courts are not a source of reliable facts. We cannot view courts as institutions that produce universal truths.[46]

In this context, Frank presents the "*Judicial Hunch Theory*." While we often view judicial rules as being completely open when approaching a case, with judges examining available facts, comparing relevant laws and statutes, and following logical reasoning to reach an objective conclusion, Frank argues this is not how judges actually work. Instead of moving forward from problem to conclusion, judges work in reverse, starting with a hypothetical conclusion, a "hunch", and working backwards. The judge finds the best arguments and statutes that support this hunch to reach their conclusion. This should not be interpreted as judges determining the conclusion at the start and ignoring everything else. Rather, the hunch serves as a starting point that guides their direction.[47]

Other research supports this view. Modak-Truran argues there is pragmatic value in this approach. The author notes that it makes their work more efficient under time pressure.[48] Furthermore, this hunch develops over time and represents a form of expertise.[49] It can be particularly useful in cases where clear rules do not provide answers. The judge develops an intuition for how cases should be decided and can more easily build arguments using analogies

[43] Llewellyn, *The Bramble Bush: on our Law and its Study*, 18–26.
[44] Marquis, Papini, Prade, Prade, Marquis and Papini, A Guided Tour of Artificial Intelligence Research: Volume III: Interfaces and Applications of Artificial Intelligence, xiii.
[45] Wacks, *Understanding jurisprudence : an introduction to legal theory*, 180.
[46] Ibid.
[47] Llewellyn, *The Bramble Bush: on our Law and its Study*, 12–24.
[48] Modak-Truran, *A Pragmatic Justification of the Judicial Hunch*, 15–16.
[49] Ibid 14–15.

and general legal principles.[50] This aligns with later research by Bladini, who argues for the value of emotions in the courtroom.[51] Emotions and intuitions are closely interconnected. Both represent non-rational influences that shape our decision-making processes. What they have in common is that they form the basis for decisions that lack direct rational foundations, as both emotions and intuitions often develop in the subconscious mind.

This perspective, that we can rely on our intuition, has faced criticism. Daniel Kahneman, in collaboration with Gary Klein, explored the complexities of intuition in their work. Klein had previously argued that we can often rely on our intuitions, but Kahneman showed that this is not always the case. They concluded that certain conditions must be met for intuition to be reliable. Intuition cannot be trusted in situations where feedback is delayed or unclear. For example, car drivers often develop good intuition because they receive immediate and clear feedback on their driving.[52]

Despite these criticisms, this may still provide an accurate descriptive picture of how judges reach their decisions. There may also be value in using emotions and intuition in legal settings that require more than just reaching a verdict. A court decision is a long process, and many small steps within that process can determine the outcome. These sub-processes may well provide clear and nearly immediate feedback, such as where the judge should direct their attention.

There are two main points in this theory within the scope of LLMs. Firstly, the theories of famous American Legal Realists such as Oliver Wendell Holmes Jr., Ferome Frank and Karl Llewellyn can bring important aspects to the field of law when incorporating LLM models. This theory can explain why we need an advanced AI system like an LLM to combine the practical field of law with artificial intelligence. This will be discussed in further detail in the section 7.1 LLM used in law. Secondly, the Hunch Theory may be very important to understand and apply when examining how LLMs work. It draws a striking parallel between the practicing lawyer and the mathematics behind the LLM. This will be further discussed in 7.1.1. The hunch theory AI-models. [53]

5.4. American Legal Realism in Sweden

A critique that might arise is why one should apply American Legal Realism to legal systems in other jurisdictions. After all, American law differs significantly from Swedish law, being based on common law versus civil law traditions. Could it not be the case that the Swedish

[50] Ibid 16–17.
[51] Bladini, *Om emotioners vara eller icke vara — ett exempel från processrätten*, 489–509. See also Richards that argue that the intuition could be a useful tool in combination with logical reasoning Richards, *When Judges Have a Hunch: Intuition and Experience in Judicial Decision-Making*.
[52] Kahneman and Klein, *Conditions for Intuitive Expertise: A Failure to Disagree*, 519–520.
[53] Chen, Fu, Yuan, Wen, Fan, Liu, Zhang, Li and Xiao, *Hallucination Detection: Robustly Discerning Reliable Answers in Large Language Models*, 123–156.

courts operate within a formalistic structure that is coherent and separate from other areas of influence?

Even if the Swedish legal system were built on a formalistic, positivist, and coherent structure of principles and rules, these would only become visible and applicable in practice through the interpretations and decisions of the highest courts. It is the highest courts, through its precedents, that creates an authoritative version of the law, where the formal structure is highlighted and clarified. For those outside the work of the highest courts, such as lower courts, practicing lawyers, or the general public, it is not possible to access these underlying principles. Thus, the law becomes apparent only through the court that has the final say in a case.

From this, one can conclude that the law is not the underlying principles and coherent structure, regardless of whether these exist or not. Instead, the law is nothing more than what the highest courts decide in their respective jurisdictions. These rulings become the face of the law in everyday life. Thus, it can be said that the law is only what judges decide, regardless of what lies behind their decisions.

This practical legal reality is further shaped to a large extent by the decisions of lower courts, particularly in cases that are never appealed to the highest court. The majority of all legal cases are resolved in these lower instances, meaning their rulings effectively become the law in practice for the parties involved.[54] While these decisions naturally adhere to the legal framework, they are also influenced by local interpretations and practical considerations that the highest courts never review or adjust. Therefore, one could argue that these lower court decisions create a legal reality that, to some extent, is autonomous from the highest courts.

Through this lens, the theory of American Legal Realism also proves relevant in Sweden, which seemingly has highly formalistic judges who follow a uniform, dogmatic method.[55] American Legal Realism therefore explains that we do not need to understand the underlying principles and theoretical models to understand the law. Just as in America, we can only understand Swedish law through the courts. There is, of course, a distinction for those who aim to develop their ability to predict how courts will rule. Those who seek to predict Swedish court decisions can more reliably build models based on what is likely to be decided by the courts. However, an internal model would have no legal value whatsoever if no Swedish court ruled in accordance with it.

This can, of course, be made more complex. The public can adhere to an idea or a "model" of the law without any court ever having acknowledged it. An example of this is the belief held

[54] This can be illustrated by the fact that in 2023, there were 286,875 cases in the Tingsrätterna (”the district courts”), 36,659 in Hovrätterna (”the courts of appeal”), and only 8,476 in Högsta domstolen (”the Supreme Court”). Similar patterns can be observed in Förvaltningsdomstolarna (”the administrative courts”) and their superior courts. *Domstolsstatistik 2023*.

[55] It is worth noting that one could argue that Swedish courts, like their American counterparts, make judgments based on their own perceptions and values. The key difference from the US is that Swedish judges share a much more uniform understanding and set of values. This is exemplified by the Swedish dogmatic method regarding the evaluation and inherent hierarchy of legal sources. Nääv and Zamboni, *Juridisk metodlära*, 21.

by many women that reporting someone for assault could constitute defamation. Although this is a myth and no court has ruled as such, it affects women by discouraging them from reporting assaults and seeking justice.[56] Should this not be considered part of the law? American Legal Realism cannot account for this and would argue that it is something other than the law.[57] This is an issue that many modern legal philosophies have addressed.

In the case of LLMs, I will conclude that the law constitutes what judges decide. There is an interesting point in this law outside the scope of the courts, and that will become evident in the final conclusion.

5.4.1. How does the judge come to a conclusion

To design an AI system that mimics a lawyer, essentially creating what American realists call "the law", one must understand how judges decide cases. The court adheres to statutory law not because it is normatively required, but simply because that is how it functions in practice, as described by American Legal Realism. For example, Günther Teubner argues that law is an autopoietic system, self-maintaining on its own.[58]

This idea is reflected in legal education: very few become professional lawyers without a law degree, and law school acts as a sieve that selects and shapes those suited for jurisprudence, who eventually become court clerks and judges.

These insights make American Legal Realism especially relevant when evaluating LLMs. Instead of trying to incorporate abstract metaphysical concepts or complex logical principles into the AI system, we can focus on the actual reasoning used by judges. We do not need to understand every theoretical principle behind the law; the key point is whether AI can replicate the outcomes that judges achieve. We only need to ensure that AI systems can mimic the results produced by judges, and in that sense, also what law students learn during their studies.

5.5. Polyvalent perspective

A critical perspective of this might involve everything else outside the court, the rules guiding everything from how we behave in a room to how the room came to be. It seems there may be many social structures outside the sphere of the courts that, in some sense, are also part of the law. The social norms defining who are considered popular, along with the environment that nudges us into particular choices, sometimes have a stronger influence on us than the court ever could. This raises uncertainty about the comprehensive nature of the law.

[56] "Myten att en polisanmälan om övergrepp är förtal drabbar kvinnor," 2021, accessed 23 November 2024, https://arbetet.se/2021/03/17/myten-att-en-polisanmalan-om-overgrepp-ar-fortal-drabbar-kvinnor/.
[57] Karl Llewellyn argues that the law's effect extends beyond individual court cases, as people in society adjust their behavior in advance based on their expectations of how courts might rule if a conflict were to arise. See Llewellyn, *The Bramble Bush: on our Law and its Study*, 23. However, this view cannot fully explain all societal adaptations since there are various reasons why people might choose not to go to court.
[58] Teubner, *Law as an Autopoietic System*.

The definition of what constitutes the law is not as clear in the polyvalent theory. It is messy, wide and open, containing elements and values from all of society.[59] This is most easily understood as the opposite of a formalistic view. The law is not an inherently logical system where all problems can, through rigorous reasoning, lead to one single solution. Due to the fact that the law contains these contradictory values, the system must accept that incoherencies are incorporated.[60]

This perspective on law, in contrast to the monocentric and purely formalistic view, becomes crucial for understanding when analyzing the role of LLMs in applying law, as will be demonstrated in 7. LLM used in law. A common criticism against AI in legal systems involves highlighting these or similar pluralistic theories to demonstrate the complexity of law. This argument holds considerable weight against those who imagine AI models requiring a purely formalistic and unified foundation.

6. The computations in LLMs

This section provides a detailed theoretical explanation of the underlying mechanisms of LLMs including transformer architecture, self-attention, and embeddings. These technical insights form the foundation for the subsequent application of LLM technology to legal reasoning, as explored in 7. LLM used in law, where we integrate these concepts with jurisprudential theories. This connection will highlight how advanced AI can potentially handle complex legal problems that extend beyond straightforward rule application.

6.1. What is "AI"?

In articles of this sort, it is very common to provide a detailed definition of "AI". However, while AI is a difficult term to define precisely, this thesis does not require an extremely precise definition. This is because we are not interested in the broad concept itself, but rather in the more specific terms like "Machine Learning AI", "Deep Learning AI", "Transformer-based AI" and "AI Chatbots", all of which fall well within most definitions of AI.

For the purposes of this thesis, I will therefore use Turner's broad definition of AI as "Artificial intelligence is the ability of non-natural entity to make choices by an evaluative process".[61]

6.2. The first AIs — Symbolic AI

To understand how modern AI systems work, it is important to understand how they do not work. Many people view AI as rule-based, pre-programmed logical machines.[62] While this is true to some extent, it mainly applies to what is called *symbolic AI*. This method is sometimes called a rule-based system. The simplest way to understand this is as a system containing

[59] Gustafsson, *Rättens polyvalens: en rättsvetenskaplig studie av sociala rättigheter och rättssäkerhet*, 27, 163.
[60] Ibid.
[61] Turner, *Robot rules : regulating artificial intelligence*, 16.
[62] For example Wahlgren in 4.2 Pre-transformers legal AI research from the field of jurisprudence.

several functions “ $f(x) =$”, where the function takes one or more inputs, performs operations on these inputs, and produces an output.[63]

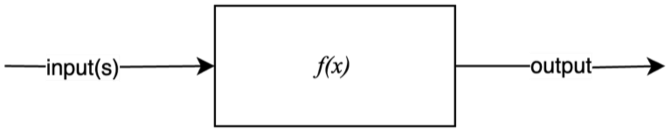


Figure 6.1 An illustration of a function "f(x)". The function takes one or more inputs, processes them, and produces an output.

For example, this function:

```python
def check_criminal(person, committed_crime):
      if person and committed_crime:
            return "criminal"
      else:
            return "not a criminal"

print(check_criminal(True, True))   # Output:"criminal"
print(check_criminal(True, False))  # Output:"not a criminal"
print(check_criminal(False, True))  # Output:"not a criminal"
print(check_criminal(False, False)) # Output:"not a criminal"
```

Figure 6.2: Python code showing a simple function “f(x)” for determining if a person is a criminal. Only when both inputs are “a person” and “committed crime” are true that the output is a criminal.

The function checks whether the input is both a person and has committed a crime. If so, the function outputs "criminal". The computer converts the program code into logical operators. The function can be much more complicated, but everything can be created with simple logical operations such as "AND", "OR", "NOT", "NOR".[64]

This very simple method is sometimes called "Good Old-Fashioned AI" (GOFAI) and can be described as "if this, then that".[65] In principle, many seemingly advanced systems can be described using this principle, from calculators to web browsers and chess computers.[66]

There are benefits to using symbolic AI. First, it is relatively easy for humans to understand and create these AI systems. To create a system with a clear purpose, a programmer can relatively easily set up which data should be the input and adjust the function to produce the desired output. Through a combination of these principles, it is possible to create systems that appear advanced to the untrained eye. Second, it is relatively easy to analyze how such a system works. For a system that produces unexpected results, the programmer can analyze what happens in the function step by step to correct the faulty decision.

---

[63] Häggström, *Tänkande maskiner*, 44.
[64] By handling numbers as binary numbers (i.e. “4” being “100”), AI can easily make mathematical operations such as addition, subtraction, division and so on. In fact, it has been logically proved that all functions can be described as a combination of NAND and NOR gates, Bird, *Engineering Mathematics*, 532.
[65] Häggström, *Tänkande maskiner*, 46, 51.
[66] ”Deep Blue,” accessed 11 nov, 2024, https://www.ibm.com/history/deep-blue.

There are also disadvantages to symbolic AI. Since it relies on very clear inputs with very clear logic, it becomes much harder to apply this to reality, which is often chaotic and unclear. A clear and classic example of this is an AI system designed to perform mathematical addition. If you input numbers that a computer system can easily understand, such as binary numbers via a terminal, an AI system can perform mathematical equations with relative ease. However, if you try to do the same thing by reading numbers written by a human, it suddenly becomes much more difficult. One can scan an image and divide all pixels into either "1" for a black pixel and "0" for a white pixel. However, the programmer will encounter problems when trying to build a function that can distinguish a 4 in different handwriting styles. Even a "4" written by the same person will differ.

Figure 6.3 A picture of me writing "4" in different shapes. A human can easily tell that all this is 4s, but nearly an impossible task for a ruled-based program.

Symbolic AI became the standard approach when creating early AI programs, mainly because programmers could directly control what they wanted the program to do. Additionally, symbolic AI requires significantly less computing power compared to non-symbolic AI. This has reinforced the understanding that this is what AI is and how it will function. When reading old speculations about AI, the emphasis on strict rules becomes evident.[67] Much of the criticism of modern AI stems from an understanding that AI works this way; A misunderstanding that AI cannot comprehend the complexities of the real world since it supposedly only works with clear logic and strict formulas.

6.2.1. ELIZA

ELIZA was an early chatbot, using symbolic AI, created between 1964 and 1967 at MIT by Joseph Weizenbaum. ELIZA was designed to imitate a psychological therapist using a relatively simple procedure. The program shares similarities with today's chatbots. Users would write messages to ELIZA, which would analyze them and respond, typically with a question.[68]

Managing therapeutic dialogue is a highly complex task, which raises the question of how an AI system created as early as the 1960s managed to accomplish this. Weizenbaum programmed ELIZA with a relatively simple task called pattern matching. The first step was to identify certain keywords. For example, if a user mentioned being "unhappy," ELIZA would respond with general questions related to these keywords, such as "Do you think coming here will help you not to be unhappy?"[69]

[67] For more, see 4. Previous research.
[68] Weizenbaum, *ELIZA: a computer program for the study of natural language communication between man and machine*.
[69] Ibid.

Despite this relatively simple process, users experienced the conversation as quite authentic, feeling as though the chatbot understood them. This AI model was not built on machine learning, neural networks, or similar technologies, but instead demonstrated an important concept: how we humans are limited in our understanding of others' comprehension. We interpret this simple process as deeper understanding. Therefore, the AI model does not demonstrate true understanding but rather imitation.[70]

6.3. Neural networks

During the early development of AI, researchers were divided into two camps: those who advocated for symbolic AI (rule-based systems) and those who supported non-symbolic AI, also known as neural networks. The idea behind neural networks was to mimic the structure and behavior of the brain, hence the term "*neural*". This approach provided the major advantage of being versatile and capable of solving complex, chaotic problems in ways symbolic systems struggled to achieve. Instead of explicitly programming rules for each decision, developers create multiple "neurons" organized into one or more layers.[71]

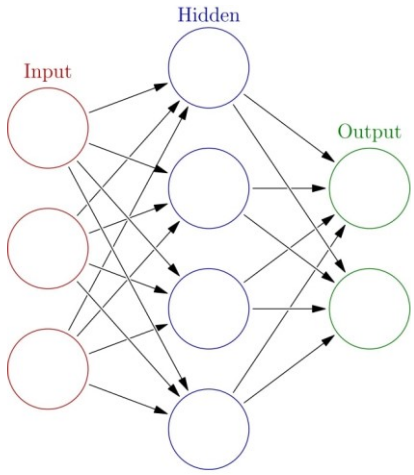


Figure 6.4 A common way of describing a neural network is as follows. The input neurons send data to neurons in a hidden layer which makes calculations to the data before sending this to the output neurons. Image by Glosser.ca, 2013.

At the input layer, neurons receive numerical values[72] (often binary, such as "1" or "0," but could also be continuous data). These input neurons forward their numerical values to the next layer, called the hidden layer, where each input is multiplied by a weight and added together. A bias term, which is essentially an additional constant added to the sum, is included to ensure that the neuron's output is not constrained to pass through a specific point.

For example, a neural network with five inputs ($X_{1\text{-}5}$):

[70] Mitchell, *Artificial intelligence : a guide for thinking humans*, 48–49.
[71] Häggström, *Tänkande maskiner*, 44–46.
[72] A numerical value is simply a number. While in AI contexts the term "values" is sometimes used alone, in this thesis "numeric values" will consistently be used when referring to numbers, while "values" alone will be used in its traditional legal sense. The latter may refer to, for example, economic or normative values.

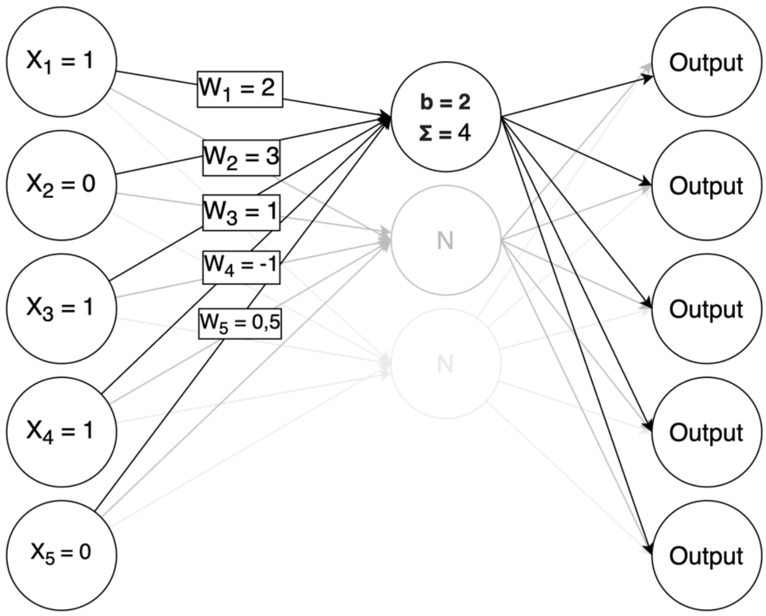


Figure 6.5 A simplified version of a neural network. The inputs $X_{1\text{–}5}$ are added with weights $W_{1\text{–}5}$ before and summed before sent to the next layer.

$$x_1 = 1,\ x_2 = 0,\ x_3 = 1,\ x_4 = 1,\ x_5 = 0$$

and a single neuron in the hidden layer. Each input is multiplied by a specific weight ($W_{1\text{–}5}$). The weights for the five inputs are:

$$w_1 = 2, \qquad w_2 = 3, \qquad w_3 = 1, \qquad w_4 = -1, \qquad w_5 = 0{,}5$$

For each input, the value of the input is multiplied by its corresponding weight, and the results are summed up:

$$(x_1 \cdot w_1) + (x_2 \cdot w_2) + (x_3 \cdot w_3) + (x_4 \cdot w_4) + (x_5 \cdot w_5)$$

$$(1 \cdot 2) + (0 \cdot 3) + (1 \cdot 1) + (1 \cdot -1) + (0 \cdot 0.5) = 2 + 0 + 1 - 1 + 0 = 2$$

After the sum of the weighted input is calculated, a bias ($b = 2$) is added. This makes the total result (z):

$$z = 2 + 2 = 4$$

This process occurs through all nodes in the hidden layers, each having their separate weights and biases. The final layer, the output layer, receives the aggregated data from all nodes in the previous layers. When a numeric value is very high, it is typically considered "activated". For example, a specific node might correspond to a particular digit in a neural network designed to recognize numbers. If the sum in the output layer is sufficiently high, the neural network interprets this as a high probability that it has identified the specific digit associated with that output node.[73]

It is very common to have several hidden layers, where the same process is repeated before going to the output layer. This method of using several "hidden" layers is often described as "*Deep Learning*".[74]

To summarize how a simple neural network works, input data is first received by nodes in the input layer. Each node in this layer then transmits its numerical values to nodes in the next layer, called the hidden layer. In the hidden layer, each incoming value is multiplied by a weight and summed together, after which a bias term is added. This sum is then passed through an activation function that determines the node's output value. This value is then forwarded to either additional hidden layers or to the output layer, where the network's final result is produced.[75]

Although the process may appear elementary (merely involving addition and summation) the interplay of multiple layers enables neural networks to tackle extremely complex problems in ways that simple rule-based systems cannot. Unlike traditional "if-then" systems, neural networks do not rely on manually defined rules for each possible scenario; instead, they learn patterns and relationships from data, allowing them to generalize and handle highly abstract concepts in a way that rigid, rule-based systems cannot. Even though modern AI systems are far more complicated than this, which will be demonstrated in 6.5 Attention, they are all in their essence built out of this principle.

6.4. Training the neural network (ML)

While it would be theoretically possible for programmers to manually set all weights within a neural network, it would be nearly impossible to achieve a network that could handle anything beyond the simplest tasks. Therefore, neural networks are typically constructed through *machine learning*, a process comparable to evolution or human learning. Initially, the weights in neural networks are set to random numbers. The network is then given a task, such as identifying numbers in images. The output of the network is compared to the correct answer, and initially, the network performs poorly since it lacks any inherent logic and merely makes random guesses. Through a process called backpropagation, the weights within the network are gradually fine-tuned to improve accuracy.

This process is repeated several times until the neural network becomes increasingly proficient at replicating the desired behavior. There are three important points to note here.

---

[73] This explanation is simplified for pedagogical purposes. For a more in dept explanation, see Nielsen, *Neural Networks and Deep Learning*.
[74] Foster, *Generative Deep Learning*, 33.
[75] For a visual understanding of this concept, I strongly recommend watching 3Blue1Brown's animation that demonstrates how a neural network identifies digits: https://3b1b-posts.us-east-1.linodeobjects.com/content/lessons/2017/neural-networks/network-propagation.mp4#t=0.001

First, this requires sufficient training material for the neural network. In the case described above, the neural network uses labeled data, meaning each image of a number is tagged with its correct value. This requires the ability to label large amounts of data.[76] When neural networks use labeled data, it is called supervised learning, as the system receives feedback when it makes mistakes. There is also unsupervised learning, which uses unlabeled data. In this case, the neural network learns to find patterns in the data independently. While this is a more complex variation, it is essentially still a form of supervised learning where the labels are indirect.

Second, AI always operates based on the material we input. This is the primary reason why bias is such a significant problem in AI. If we input biased training data, the neural network will inevitably produce biased results.

Third and most importantly, AI developers do not need to understand the internal logic that forms when neural networks are trained. The methods and weights that the neural network develops appear completely random to humans. This means it can be very difficult or nearly impossible for developers to control the results except through the selection of training data.

The learning process is an enormously important part of AI. While I will not dive as deeply into this topic going forward, focusing instead on AI functionality, it remains crucial knowledge for anyone developing or working with AI.

6.4.1. Overfitting

Another important concept regarding training is called overfitting. In simple terms, it means that an AI has been trained so narrowly that it finds specific solutions to many different problems rather than developing a general approach that can be applied to multiple cases. A good example of this is someone trying to become skilled at chess. Instead of learning how to think and understand the logic behind chess, the person memorizes specific moves and strategies. This player will perform extremely well when the opponent follows the moves and strategies they have learned. However, if the opponent makes a move that the person has not trained for, they will find it very difficult to evaluate the situation. This happens because the person has overfitted their training. They have been trained to handle a specific game excellently but have not learned to generalize, making them significantly worse at handling situations they have not trained for before.

This becomes crucial in AI because we almost always want AI to be able to handle cases it has not been trained on, using principles learned from the training data.

[76] See for example Google using the input data from reCAPTCHA (the method webpages to confirm you are a human) as classified training data for AI ”Google reCAPCHA, Creation of Value,” 2024, accessed 2 December, 2024, https://www.google.com/recaptcha/intro/?zbcode=inc5000#creation-of-value.

6.4.2. Generators

Neural networks are generally divided into two separate types. Previously the type that can identify something has been described, for example identifying what number a picture represents. These types are generally known as classifiers, since they classify an input.

The other main type that this thesis will focus most on, is the neural networks known as generators. While the technique behind them is mostly the same, the purpose of the neural network is completely different. The generators have the main purpose of creating, generating something, usually from an instruction, better known as input or prompt.[77] Early developments of this were text-to-image generators with the task of creating an image from a text input. There are also text-to-text generators with the task of, often, continuing a text from an input-text. This type of format to format can be created in all kinds of ways and an AI-model which can handle multiple formats is usually called multi-modal.

an image of a law student studying for a test	
John wanted to thank his neighbor by giving him a big bottle of_	_his favorite wine. After all, his neighbor had gone above and beyond by helping him fix the fence that had been damaged…

Table 6.2 Example of text-to-image (top) and text-to-text (bottom). This has been generated through ChatGPT-4o.

While the process of creating images from text requires deeper understanding, text generators are considerably simpler to grasp and share more similarities with image classifiers. Instead of classifying images, text generators attempt to predict the most probable next word in a sentence. Given a text (prompt), the text generator predicts the most likely next word. This new word, combined with the previous input, is then fed back into the generator to predict the next word in the sequence, and so on.[78]

---

[77] Kronblad, Pregmark and Berggren, *Difficulties to digitalize: ambidexterity challenges in law firms*, 42, Rogers, *A first course in machine learning*.
[78] S, Kadry, D and Aruchamy, *Generative AI and LLMs: Natural Language Processing and Generative Adversarial Networks*, 53–55.

While there are many interesting aspects of formats other than text, this thesis will focus on text-to-text generators, as this is where most of the actual heavy processing for handling law will occur. However, it is possible that multi-modal AI might be necessary for AI to reach its full potential in handling legal matters.

6.5. Attention

The previous examples of neural networks are often called Convolutional Neural Networks (CNNs).[79] These have shown spectacular results, as demonstrated by projects like Intel's DeepMind[80]. Various iterations have been developed, such as Recurrent Neural Networks (RNNs)[81] and Generative Adversarial Networks (GANs)[82]. While each of these variations has shown potential and contributed to AI evolution, they all demonstrated limitations when applied to increasingly complex areas.

None of these earlier neural networks were particularly suitable for the vast and complex field of law. While some could handle specific legal tasks, it was not until the emergence of the attention mechanism that neural networks could manage more complex situations effectively.

The attention mechanism is a key component of what we call *transformers*. To understand this aspect of AI, one needs to examine the underlying mathematics. Although this can be quite complex, especially for practicing lawyers without a mathematical background, it is essential to understand how these AI models handle legal matters and why they are more competent at doing so than any previous AI systems.[83]

This self-attention mechanism allows the AI to weigh the importance of various elements in a legal text, which is crucial for interpreting ambiguous or complex legal scenarios that traditional rule-based systems lack.

6.5.1. Tokens

Before exploring the complexities of the transformer, it is essential to understand how these AI models process input. While image-based AI models typically split input into pixels, text-based models need to divide text input differently. Although splitting text into words might seem like a natural approach, modern AI models prefer to break words into smaller components. This is precisely what all modern chatbots do — they divide words into subparts called "*tokens*".[84]

---

[79] Wang, Turko, Shaikh, Park, Das, Hohman, Kahng and Polo Chau, *CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization*.
[80] Mnih, Kavukcuoglu, Silver, Graves, Antonoglou, Wierstra and Riedmiller, *Playing Atari with Deep Reinforcement Learning*.
[81] Tealab, *Time series forecasting using artificial neural networks methodologies: A systematic review*.
[82] Noll, AI, digitalisering och rätten : en lärobok, 227.
[83] S, Kadry, D and Aruchamy, *Generative AI and LLMs: Natural Language Processing and Generative Adversarial Networks*, 50–51.
[84] Ibid 102.

```
Short words are often a single token, while long, complicated sesquiped
alian words are often split into several tokens. The same is often true
for missplelings.
```

Figure 6.3 Image of how ChatGPT tokenize a sentence. Each colored block is a separate token.[85]

```
[15900, 6391, 553, 4783, 261, 4590, 6602, 11, 2049, 1701, 11, 25334,
6707, 38582, 295, 175486, 6391, 553, 4783, 12648, 1511, 4919, 20290, 13,
623, 2684, 382, 4783, 1343, 395, 5141, 789, 45539, 364]
```

Figure 6.4 Each token is stored as a number. This is the numbered tokens of the one from the previous image. Notice that the same token has the same number, for example "often" = 4783.

Long words may be split into several tokens while shorter words often remain as a single token. This is the "data" that modern AI systems with transformers work with. This represents the "word dictionary" of a modern AI system. For the sake of simplicity, the rest of this thesis will treat every word as one token.

6.5.2. Embeddings and training of transformers

As previously described, all words (tokens) that an AI system can handle are stored in a dictionary within the AI system. Rather than describing and defining each word individually, they are placed in a multidimensional space called an embedding matrix.

This may be misinterpreted as the AI system operating beyond 3D space or understanding higher dimensions. However, these embeddings are simpler than they may appear. As we will see later, these dimensions may rather be interpreted simply as different scales of various properties.

For example, one dimension might represent "*size*", another "*color*", and another "*number of pages*", and so on. These dimensions can represent any property. When visualizing these embeddings, we typically plot two, and sometimes three of these properties as dimensions to illustrate how an embedding look. This will also be done in this thesis. While I will use two dimensions for illustration purposes, embeddings can be created with any number of dimensions.

All words (tokens) within modern AI have a unique position within the embedding. For example, the words "Stockholm" and "Gothenburg" each occupy their own distinct place within this embedding.

[85] https://platform.openai.com/tokenizer

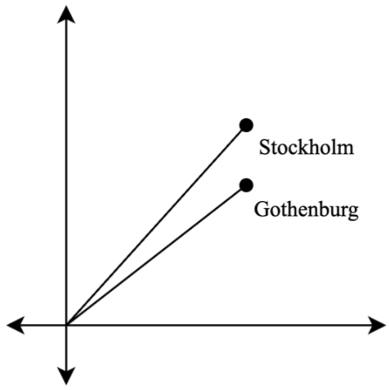


Figure 6.5: A vector space where the words "Stockholm" and "Gothenburg" are represented.

In this example, we have two dimensions where Stockholm and Gothenburg are both represented by coordinates in this embedding. They are fairly close to each other along one axis since they are both cities within Sweden (their properties are similar). In embeddings, you will find similar words positioned close to each other. The other dimension beyond "big city in Sweden" may describe different properties (such as "*west/east*", "*south/north*", "*population*", or even "*number of football teams*"), which explains why Gothenburg and Stockholm are slightly separated.

The fascinating and important aspect for this thesis emerges when we analyze the distances from "Stockholm". If we draw a line from Stockholm to another word, such as "Gröna Lund", and then draw a line of the same length and angle from Gothenburg, we arrive at a point very close to Liseberg. This is not coincidental, through rigorous machine learning, the AI system has learned the relationship between these pairs. This can be understood as lines going in the same direction representing an "*amusement park*" property. Gröna Lund has mostly the same relationship to Stockholm as Liseberg does to Gothenburg.

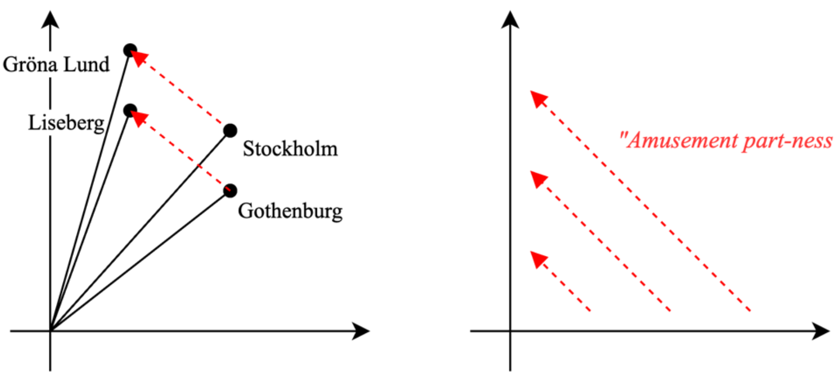


Figure 6.6: A vector representation showing the relationship between cities and their amusement parks.

LLMs store all words as vectors with coordinates, for example, $Stockholm = [5, 6]$ and $Gothenburg = [5, 5]$, rather than representing them visually. This makes it easy to scale since adding a dimension simply requires adding another coordinate. Three coordinates would represent a word-vector in 3D-space, and n coordinates represent a vector in n-dimensional space. For example: $Word = [c_{1,}, c_2, c_3, \ldots c_n]$. While it is difficult to visualize vectors in spaces with more than three dimensions, it is easier to think of dimensions as different property ranges (for example "*blueness*", "*year*" but also "*amusement part-ness*" and all kinds of abstracts). The coordinate of a word vector therefore specifies the property of that word. It is important to note that while vectors in multi-dimensional spaces are difficult to visualize, mathematical principles remain largely the same. Calculating the distance and angle between two points in 3-dimensional space is fundamentally similar to performing these calculations between points in, for example, a 7-dimensional space.

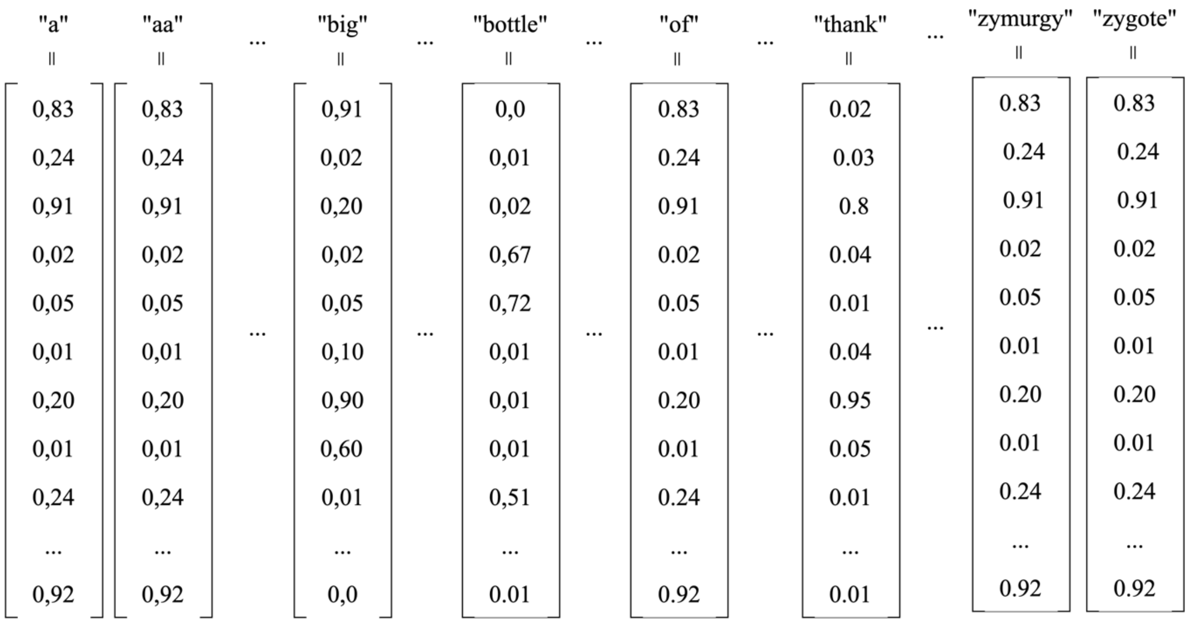


Figure 6.4 This image shows how an embedding matrix can be structured. It can be compared to a dictionary, where each word has its own vector with different numerical values in each dimension. These numerical values may symbolize various properties of the words.

It is important to note that the AI system does not have any inherent understanding of each word. Words are merely represented as numbers in an n-dimensional space, and their meaning is defined by their distance to other words. The directions in the embeddings can therefore be seen as describing meaning to words in relation to other words. Therefore, the more dimensions an embedding contains, the more ways an AI system may categorize and systematize all the words. This may make it seem that there can never be more "properties" than there are dimensions. This is true for embeddings with few dimensions; however, in embeddings with a high number of dimensions, it is actually possible to create dimensions with a greater number of "properties" than there are dimensions through something known as superpositions.[86]

This understanding of how AI systems comprehend words (tokens) implies that all words and meanings exist only in relation to other elements. Words, phrases, and the way language is

[86] "Toy Models of Superposition," 2022, accessed 29 december 2024, https://transformer-circuits.pub/2022/toy_model/index.html.

constructed influence these words and can give them entirely new meanings. This can be compared to Wittgenstein's later work where he describes how words and phrases are determined through "language games".[87]

6.6. Transformers and self-attention

Transformer architecture gained popularity after the publication of the article "Attention Is All You Need" in 2017.[88] The attention mechanism itself was initially introduced to improve sequence-to-sequence (seq2seq) models, addressing inefficiencies in handling long sequences. Previous AI models like RNNs (Recurrent Neural Networks) processed input sequentially, one word at a time, which limited their ability to retain important information from earlier parts of a sequence. This made RNNs poorly suited for tasks involving long texts, as they struggled with both efficiency and output relevance. Transformers overcame these limitations by replacing RNNs with a self-attention mechanism, enabling the model to focus on all parts of the input sequence simultaneously and process data in parallel, making both training and inference significantly more effective.[89]

Transformers brought a revolutionary change to AI, significantly improving the learning process of AI systems by centralizing the self-attention mechanism.[90] This mechanism allows transformers to efficiently process input and capture relationships within data, making them far superior to previous AI models like RNNs.

When a modern AI system processes text, each word or token is initially converted into a vector representation using an embedding matrix, which encodes the word's properties based on prior training. However, static embedding alone is insufficient to understand the meaning of a sentence, as they lack contextual relationships between words.

To address this, transformers use the *self-attention mechanism*, which enables the model to dynamically adjust the representation of each word by analyzing its relationship to every other word in the input. For example, consider these three sentences using the word "*well*":

> *"The villagers drew water from the ancient well every morning."*
> *"She performed well on her final exams."*
> *"Well, I suppose we could try it another way."*

These examples demonstrate how the same word can have completely different meanings depending on its context. For the AI system to understand which meaning of *"well"* is being used, it examines the relationship between *"well"* and the other words in each sentence. When *"well"* appears at the start of a sentence followed by a comma, it typically functions as a conversational filler or expression of thought. When preceded by "performed," it serves as an

[87] Wittgenstein, *Filosofiska undersökningar*, 40–51.
[88] Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser and Polosukhin, *Attention Is All You Need*.
[89] S, Kadry, D and Aruchamy, *Generative AI and LLMs: Natural Language Processing and Generative Adversarial Networks*, 234. Foster, *Generative Deep Learning*, 206–208.
[90] Towes, *Transformers Revolutionized AI. What Will Replace Them?*

adverb meaning "good." The self-attention mechanism checks each word against all other words in the input to determine how context affects meaning.

Take this example:

"John wanted to thank his neighbor by giving him a big bottle of_"

"John wanted to thank his neighbor by giving him a big bottle of_"

The interpreter of this sentence needs to interpret several words by their context compared to other words. Who wants to "*thank*," who wants to "*give*" needs to be determined. First, the word "*his*" is a reference to John, then "*him*" is a reference to his neighbor. The word "*big*" will affect the next predicted word. The list of all the rules of how each sentence in this world is supposed to be affected can be endless. Furthermore, another sentence may be constructed completely differently. Building an AI system with all these inherent rules would be impossible. That is why it would take a system such as the one with the transformers with its attention-mechanism to be able to understand the context and meaning of words.

To illustrate how this works, consider an analogy of a blind person in a room. In this room, they have arranged all their necessary belongings. When they need to retrieve a ball, they cannot see where it is, but they know they need to walk two meters north and four meters west to reach it (coordinate [2, 4]). Similarly, they know a book is located four meters north and five meters west (coordinate [4, 5]).

Sometimes, they might need to find a specific item, like a blue ball. Although they cannot see which ball is blue, they know that blue items are typically placed about 20 centimeters northwest of the standard object. Therefore, they can find a blue ball by first going to the standard ball's location and then moving 20 centimeters northwest.

Similarly, they know that to reach the 17th-century books, they need to move 30 centimeters north from the standard book location, and from there, 10 centimeters east to reach the novels. In this way, the blind person can be confident they are picking up Don Quixote without being able to see the cover.

Likewise, a large language model can apprehend numerous aspects of our world without ever directly experiencing it. Instead, by meticulously analyzing vast repositories of text, it skillfully simulates understanding, capturing the subtle nuances and complexities of human knowledge and perception across disciplines.

6.7. Key takeaways

In this section, I have outlined the key technical elements that underpin modern LLMs. I have examined the transformer architecture, with a particular focus on the self-attention mechanism, tokenization, and embeddings. These components enable LLMs to convert raw text into multidimensional numerical representations and dynamically weigh contextual relationships between words. This dynamic processing is fundamental for capturing the subtle nuances and complexities of language.

Understanding these technical principles is essential for the next phase of my research. In, 7. LLM used in law, these mechanisms will be applied to a legal context. Specifically, I will explore how the ability of LLMs to interpret and manage complex linguistic patterns can be leveraged to analyze legal texts and predict judicial reasoning. By linking the technical foundation established here with jurisprudential theories, I aim to demonstrate how advanced AI models can address intricate legal problems that extend beyond mere rule application.

The established technical framework provides a solid foundation for exploring the practical implications of LLMs in legal reasoning in the next section.

7. LLM used in law

This section is the so-called "techno-legal analysis" where I analyze the theoretical capacity of LLMs to solve legal problems. Here, the theoretical foundation for law established in 5. Theoretical framework is applied to the theory of how LLMs function as established in 6. The computations in large language models.

Now that it has been established that modern AI models built on the attention mechanism via a transformer model can create functions resembling human reasoning, it is interesting to investigate whether this can be applied in the limited field of law described in 2. Problem and purpose.

Before examining this topic further, it is worth noting that XAI (Explainable AI) will essentially be discussed in parallel. Legal reasoning largely involves demonstrating and arguing why a particular solution is the most applicable in a specific case. Since this includes developing one's viewpoints, it essentially reveals how the AI system reached its conclusion.[91] This becomes even more evident in the next section[92], where the chatbot must explicitly show its reasoning processes.

[91] This can also be divided into "process of discovery" and "process of justification." The latter most closely resembles judicial reasoning, and I argue that the same applies to these AI systems. An LLM constructs sentences in its justifications, but the exact discovery process remains more difficult to observe as it occurs within the mathematical algorithms. See Strömholm, *Rätt, rättskällor och rättstillämpning : en lärobok i allmän rättslära*, 424–425.

[92] 8. Real world test of LLM in legal reasoning

7.1. The formalistic view versus the non-formalistic view

Early versions of the AI evolution showed prominence in symbolic AI. I will demonstrate that this has a close connection with the formalistic view of the law. Just as the purely formalistic view of law treats it as a self-contained, predictable, and uniform system; a symbolic AI operates in a similar manner. When dealing with a complex situation, a highly intelligent lawyer would demonstrate the most logical way of solving the question purely from within the formalistic school of law. The lawyer would analyze all sources and put forward a formula or flowchart that leads the problem to the one correct solution. Any issues along the way would be solved within the system. Every requisite would have one correct definition, and every consideration would be weighed properly. The symbolic AI works in the same manner as the highly intelligent lawyer. The symbolic AI has a purely logical way of solving the question. The symbolic AI uses all available legal texts, which are neatly sorted and integrated into its algorithm. Using this procedure, it produces a consistent answer that follows all the rules within all the legal texts. Any issues along the way are solved by internal algorithms, all requisites and terms have their own definition that it uses, and mathematical algorithms determine the appropriate weightings in each case.

As demonstrated by theories of Legal Realism and polyvalent perspectives, it is problematic and impractical to see the law as an enclosed system. The law is deeply connected to every other aspect of our world. The law cannot be defined in isolation but must be understood in relation to the broader context. Furthermore, the law does not produce purely logical conclusions that follow a strict path to a single solution. Instead, it is complex and dependent on factors previously considered outside the scope of legal consideration. Solutions may vary significantly depending on location, circumstances, and perspective. All attempts to fully formalize the law have proven unsuccessful. Just as mathematics cannot prove its own consistency purely from within, no legal system exists in isolation from the outside world.[93]

Given that the law is not fully formalistic, it remains uncertain how a symbolic AI for legal applications might be effectively created. There is considerable ambiguity in designing algorithms that produce varied results when applied to real-world objects, especially in situations with unclear definitions and ambiguous boundaries. Moreover, it remains uncertain how an AI system can process and organize legal sources when any source might be legally relevant without a clear hierarchical order.

For those who view AI as a symbolic system, a robot following strict algorithms, legal AI appears impossible. However, this limitation applies only to symbolic AI. It would be a mistake to assume the same constraints apply to non-symbolic AI.

Non-symbolic AI does not rely on strict formulas; its internal decision-making processes are complex and fluid. Rather than defining objects through single, true definitions, this type of AI system perceives objects as points in a multidimensional space, comparing them to other

[93] See section 5. Theoretical framework

objects. These objects are defined through millions of sources, and their meaning can vary significantly depending on context.

Consequently, in a post-formalistic legal world, non-symbolic AI may have greater potential in the field of law than previously thought.

7.1.1. The Hunch Theory in LLM-models

The Hunch Theory could be understood as the intuition of the lawyer. When a new case is presented, the lawyer relies on intuition to decide on a solution instead of following a strictly logical, step-by-step process to reach a conclusion. Examining modern AI-systems, especially those created with the attention-mechanism, a very similar phenomenon of the hunch theory starts to emerge. While it is true that the AI-systems with an attention mechanism still only predict the next word in sequence, I argue this is miss viewing and the AI systems are, in fact, creating a hunch much similar to a layer predicting a word. Take this example:

> *“A man cracked the window and cleared the glass. He reached into the window and slowly turned the lock_”*

Making ChatGPT continue this sentence creates this continuation of the meaning:

> *“_on the inside of the door.”*

The next word ChatGPT predicted was the word "*on*", but if the AI only was to predict the next word, as in one task, it would probably predict the word "*open*", as several words hint that this is the case. But the AI does not immediately predict the word "*open*", instead it predicts the word "*on*". My interpretation is that the AI is close to predicting the word "*door*". To make "door" fit in, it makes up a sentence so this would fit, hence it fills in the words "*on the inside of the*" before "*door*".

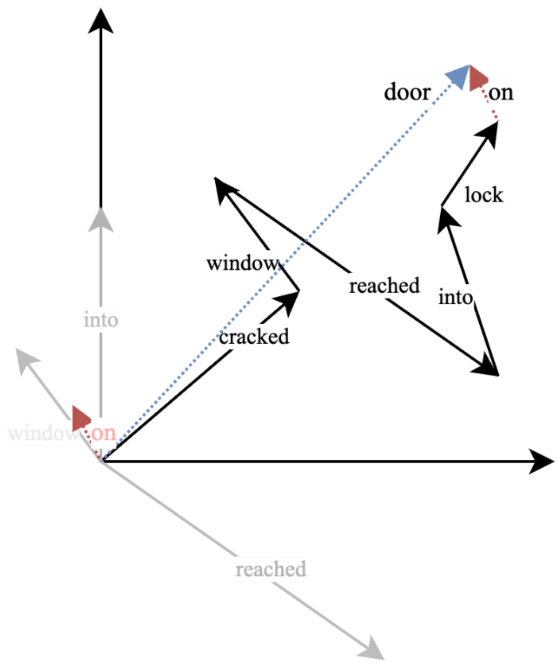


Figure 7.1 This image demonstrates how the words in the example above (showing only the most important words, actually tokens) cause the model to "walk" closer and closer to "door". At the end, the word "on" is added to arrive at almost precisely "door". In reality, these consist of many more dimensions than just two, and there are additional mechanisms for how words are selected. However, this represents the essence of how it works.

Just like humans getting a hunch or having an intuition for something further than just the next word, the AI predicts something likely further along and then tries to build a logical way of reaching there. Of course, all these steps are calculated where all the individual words are used by the choice of the highest likelihood of being there. During the attention-mechanism, the AI understands that several of the words together are assembled to lead the sentence to a "door". The AI-system knows though, that meaning cannot be constructed as *"...slowly turned the lock door*", but there needs to be some other words in between to make the words sound and follow the rules of how sentences are constructed. It therefore adds the words that are most predictable to bind the initial sentence and the word "door". The chatbot follows this process for every word it predicts, and continues with:

> *"With a faint click, the lock gave way."*

This is, in my view, very similar to the hunch and the intuition of the lawyer. The lawyer sees a problem and an intuitive future in how this can be solved. The lawyer then finds the meaning and the rules to be set in between to make the sentence logical and sound.

This should not be interpreted as if lawyers only work with their intuition and only find the middle ground to fit the intuitive conclusion. Very often the lawyer finds that the logical bridge between the problem and the solution may be impossible. For example, the intended solution for a problem may need a rule to be fulfilled, but the lawyer realizes along the way that only three out of four of these requisites are fulfilled. The lawyer then creates a new intuition and finds a new path to the solution. This continues until the lawyer has created a fully logical and sound solution. This can be described as follows:

1. *The initial problem with all the circumstances.*

The lawyer now finds a hunch that this problem should be solved in a certain way

2. *The initial problem with all the circumstances. → (intuition) → Solved in a certain way "S_1".*

In the next step, the lawyer tries to find the logical solution to how this is solved in the "S_1". This is also brought up by intuition.

3. *The initial problem with all the circumstances. → Rule "R_1"? → Solved in a certain way S_1.*

The lawyer then attempts to construct a logical bridge between the problem and the solution using rule "R_1". During this process, the lawyer may discover that not all prerequisites of rule R_1 are satisfied. As a result, the initial solution S_1 is discarded and replaced with a new potential solution S_2.

4. *The initial problem with all the circumstances. → (intuition) → Solved in a certain way $\cancel{S_1}$ S_2.*

The lawyer then follows the same process with solution S2, which may require a different rule "R2". In this case, all prerequisites of the rule are satisfied.

5. *The initial problem with all the circumstances. → Rule "R_2" → Solved in a certain way S_2.*

The AI works in very similar ways, initially. The combination of the words in a problem may lead to the probability of the next word to point in certain directions. A set of circumstances may similarly point to the solution of "S_1"; the AI then puts the most probable words in between these two to bridge them. Just like the lawyer, the road from the problem with all the circumstances may be impossible. The AI model will then find these in the middle of the bridge, since these "bridge" words also will affect the solution, and they may steer away from "S_1".[94]

Since chatbots cannot change their course midway, this becomes a problem. The AI model will start going to a solution, for example by using R1, as it initially seemed likely to lead to a solution. When the AI model "discovers" that not all prerequisites are met, issues arise. It cannot delete what it has already written. Instead, it will incorporate its previously predicted words as part of the solution and try to solve the problem with them. This causes LLMs to

[94] Compare with this article which has examined how future words influence language models during prediction: Xie, Xunying, Yu, Ragni, Wong and Gales, *Exploiting Future Word Contexts in Neural Network Language Models for Speech Recognition*.

make errors, especially when creating solutions that require multiple sequential steps. This means that LLMs have the ability to solve legal problems that can be addressed through their “intuition”. Just as a more experienced lawyer is often correct when following their intuition, increasingly trained LLMs become better and better at solving legal problems.

7.2. The non-intuitive problems

Several legal problems obviously cannot be solved through intuition alone. They require a logical step-by-step analysis to reach a conclusion. Consider a case where a storage company's facilities have been affected by mold, causing damage to customers' belongings. Few lawyers could solve this intuitively. Instead, the lawyer needs to methodically examine the circumstances and approach the problem logically. First, the lawyer needs to determine which law applies, which rules within this law govern the situation, whether the parties have any private agreements, and if these agreements take precedence in certain aspects. Additionally, the lawyer must investigate if there are any liability provisions that address this, and once established, review what costs a party is entitled to. The lawyer would find it very difficult to intuitively determine the final question without first solving it step by step.

This is where LLMs begin to struggle. Even though an LLM has multiple transformer layers, it cannot internally solve all these steps statistically. This is because the problem needs to be solved *sequentially*, where the conclusion from the first step affects the application in later steps. However, the AI model processes everything simultaneously and analyses what is statistically most likely based on all available information. It cannot employ purely logical argumentation where, for example, the first step logically should have significancy. This becomes evident when asking AI complex questions.

For example, there may be several factors suggesting that a contract should be interpreted as a specific type of contract where special regulations apply. However, even if 90% of the factors indicate that the contract should be interpreted this way, the remaining 10% might include a requisite that is not fulfilled, meaning the regulation in question cannot be applied. An AI model that intuitively begins to apply this regulation will continue to apply rules from that regulatory framework, leading itself further and further astray in its application. This results in an outcome far different from what an experienced lawyer would produce in the same case.

7.2.1. Chain of Thought (CoT)

There is a solution to the problem that cannot be solved only through intuition. By breaking down complex legal problems into smaller components, they become significantly easier to solve. A legal professional understands that law primarily deals with relationships, so a natural first step is to break down legal problems into their relevant relationships. Even though legal cases may involve multiple parties and actors, it is almost always possible to break down the problems into relationships and handle them step by step.

The next step is determining the specific issue between two parties. Once this is established, the lawyer can begin searching for sources that address the specific problem. The lawyer may then look for relevant legislation, agreements between parties, or established legal precedents. After finding these sources, the lawyer matches these laws, agreements, and precedents against the circumstances of the current case, verifying whether the described legal applicability actually matches the situation at hand. After this, lawyers examine what sub-problems this relationship contains, and so on. Whenever a problem arises, the lawyer can break it down into smaller components to handle them.

A lawyer who follows this process will, after sufficient subdivisions, be left with sub-problems where the "logic" is not far removed from pure linguistic questions. For instance, in determining whether an agreement constitutes rental or storage premises, one might find that storage law requires "*the company to assume a duty of care or custody*." The linguistic comparison of circumstances, interpreting whether the words used in the question align with "*a duty of care or custody*", becomes a matter of determining which interpretation is closest. What becomes apparent here is that one arrives at a linguistic question of how to interpret words, precisely what LLMs excel at. This type of breakdown can be applied to most types of logical problems where you ultimately end up with linguistic considerations. This means that LLM systems facing a complex problem must break down their problem into smaller components to use their expertise in language to solve legal problems.

Just like one part of being a lawyer is to figure out the method of splitting up a legal problem into smaller steps. The AI model can similarly split it into steps that it itself would have to solve. The chatbot works with itself, first deciding how to split the issue into smaller parts, then feeding the output to itself to solve it in this way. This is what newer chat models that are stated to be able to reason do. This process is often referred to as *Chain of Thought* (CoT).

7.3. Testing LLM in practice

LMM is a rapidly evolving field. At the time of writing, the most popular recent LLMs are OpenAI's ChatGPT-4, Google's Gemini 2.0, and Claude 3.5 Sonnet. Based on the theoretical background and conclusions, I have conducted several tests to see if theory aligns with practice. Testing AI models is challenging for two main reasons: first, different prompts and examples can yield varying results. Second, it is difficult to determine whether an AI model is solving a new problem it has never encountered before, or if it has been trained on similar information previously. Therefore, I have attempted to create new examples to the greatest extent possible.

While these tests were conducted using all three AI models mentioned above, only the results from OpenAI will be demonstrated, as it is the only model that allows for sharing conversations. It is also the most well-known model.

My experiments show that chatbots demonstrate good proficiency in applying basic legal questions when asked to respond as a legal professional. The AI models understand that they should solve questions by applying Swedish legal regulations. They appear capable of accessing Swedish regulations without difficulty (they are either trained on Swedish legislation, able to access Swedish legislation through an API, or have the legislation as an

RAG. Most probably the first).[95] This demonstrates that LLMs, using their attention mechanisms, can perform what is central to legal application. That is, taking circumstances, finding a relevant rule, and then applying the circumstances to reach a conclusion. This is not particularly surprising, as Swedish legislation can be viewed as instructions for chatbots. When given circumstances, chatbots can use the attention mechanism in transformers to combine the question with applicable law, similar to natural language, to arrive at a suitable conclusion.

In other words, the chatbot uses the semantics and syntax of the legal question to identify which legal rule is applicable. It accomplishes this by using vectors in the embedding matrix. Once the relevant law is found, it uses the words in that law, along with its syntax and semantics, combined with the legal question, to calculate the most probable solution. The chatbot thus first "walks" to the most appropriate legal rule, and from there “walks” to the most suitable legal conclusion. One could say that the chatbot solves these steps intuitively through its understanding of language.

As long as there is a rule that addresses a situation, chatbots should have no problem arriving at a suitable solution. This can be called literal interpretation, and since chatbots excel at natural language processing, they will be able to reach appropriate solutions. Even problems that are difficult for humans, such as old or complex Swedish texts, pose no comprehension challenges for chatbots.[96]

There are many other methods for solving legal problems besides literal interpretations. One example of this is analogical reasoning. One could describe analogies as rules that match about 80% of the circumstances, but since there is no direct rule in the area, the closest or most similar regulation must be chosen instead. This is similar to comparative law, where other legal areas or jurisdictions are examined for inspiration. In both cases, it could be described as investigating whether there is an underlying rule that can be discerned by examining other rules and then applying this underlying rule.

Since LLMs do not need exact rules for predicting the next word, this is not a problem. Just as language is full of complex and diffuse rules that humans master without knowing exactly why, LLMs are designed to understand these patterns. Therefore, they have no difficulty learning and understanding these legal methods through training on data where these methods have been applied.

I wanted to investigate the extent to which LLMs can understand analogies in domains that are not inherently legal in nature. In other words, it remains uncertain whether AI models truly

[95] See examples of conversations with ChatGPT: https://chatgpt.com/share/67910033-1400-8004-b5b1-a3e38a2488eb, https://chatgpt.com/share/6791017f-5f38-8004-b45a-371f3feea00a, https://chatgpt.com/share/6791021d-b898-8004-9d6d-b8152ace5788, https://chatgpt.com/share/679103aa-80d0-8004-9dfb-5c0c7ee14356

[96] For example, consider the Promissory Notes Act ("Lag (1915:218) om avtal och andra rättshandlingar på förmögenhetsrättens område"), which was written more than 100 years ago. While students often find its language difficult to comprehend, chatbots have no trouble understanding it. https://chatgpt.com/share/67920b35-128c-8004-8e07-8b87a4651232

grasp the concept of analogies or if they have simply been trained to recognize situations where analogies apply. To test this, I created a fictional example involving an unusual sport where there was no explicit rule. Despite this lack of direct regulation, the LLM was able to identify and apply rules from the most closely related scenarios.[97] This demonstrates that the concept of analogical reasoning is inherent in these AI models.

Since the dimensions in LLMs do not need to be based on clearly defined functions, or even be interpretable by humans, they can build complex abstract relationships. It would therefore be naive to suggest that AI models cannot understand our morals or ethics. According to my theory, this does not mean that the models need to possess any real morality or ethics (whatever those may be) they only need to be able to mimic humans. LLMs that can predict ethics similar to how a judge would express them therefore possess, in that sense, morality and ethics. Moral considerations that should take precedence over the application of other rules are thus very possible for an LLM to learn. These moral considerations do not even need to be clearly defined; they can even contain contradictions, and these can still be created within an LLM's network.

When testing AI models on technically legal but morally reprehensible actions, the AI model is clear about this:

> *"While it's legally possible to send a cease-and-desist letter, doing so without a strong basis can have significant ethical, legal, and reputational consequences."*[98]

LLMs are therefore theoretically well-suited to handle a polyvalent legal structure (perhaps even better than a purely formalistic one) where contradictions, vague and abstract values influence the law, and where one-sided answers do not exist.

It is in fact, more difficult to design concentrated questions that in theory is impossible for LLM to handle. Gärdefors attempts to demonstrate that AI struggles with tasks requiring world understanding.[99] However, his examples, which often involve physical world scenarios, are difficult to translate into legal problems. Moreover, these examples are challenging even for humans to solve.

One way to "trick" AI models is to use literal interpretations of words that often carry implicit meanings in everyday conversation. Since AI models are trained in human language patterns where we use words with their implied meanings, there is a high likelihood that AI models will interpret these words in the same contextual way rather than literally.[100] However, these types

---

[97] https://chatgpt.com/share/6790fb95-12ec-8004-b8e1-d694505dbc95
[98] https://chatgpt.com/share/6752d011-90f4-8004-aa8f-d1e2ecd71f8a
[99] Consider the example of people crossing a bridge. It takes time for humans to arrive at a solution, and some people would likely struggle to find one or make the same mistakes as AI models do. It is also possible that future AI models will be trained on more data that gives them a clearer understanding of gravity when applying solutions to "real world" scenarios. Gärdenfors, *Kan AI tänka? : om människor, djur och robotar*, 108.
[100] In one example, I exploit the fact that when describing "one half," there is often an underlying implied meaning that the other half should be different. https://chatgpt.com/share/6792289b-a2e4-8004-9668-b91a9cea7250

of questions often "trick" humans as well, and according to the theoretical background, AI performance must be compared against human performance.

This demonstrates how difficult it is to conceive legal methodology examples that would be impossible for an LLM to handle. If a method exists, there is typically a way for an LLM to learn it. The examples where LLMs fail to perform become so hypothetical that even humans struggle to handle them.

However, there is a clear challenge for AI models: handling long, complex questions with ambiguous problem statements. This becomes more evident in the examples shown in an example.[101] As more rules, people, and contexts are introduced, the risk of LLMs making mistakes increases rapidly. These errors can be significant, as a small error at the beginning of a legal response will influence subsequent conclusions in the text. This is especially problematic for LLMs as they continuously build upon what they have previously written; LLMs have no ability to go back and revise their answers. The chance of errors appearing in the response increases and affects the answer. Furthermore, LLMs struggle to understand what specific aspects I, as the questioner, am interested in for each case. AI models rarely present and evaluate multiple possible scenarios.

When experimenting with this approach, I found that the chatbot produced the most accurate answers when dividing the questions into five parts, similar to how a lawyer would structure the analysis in the example above. These five parts were presented as separate prompts, each including examples but without explicitly stating how the specific judicial problem should be answered.

My research shows that the complexity for LLMs does not lie, as many previous legal researchers suggest, in non-formalistic questions. This was a problem for symbolic AI, not for LLMs. In this case, one can imagine that formalistic problems are issues where most people (trained in the field) arrive at the same conclusion, with pure mathematical conclusions being the most extreme example. On the other hand, there are problems where most people (trained in the field) could arrive at different conclusions. Extreme examples here are purely ethical and moral questions, and legal problems that have no nearby rules whatsoever.

However, there is a mistake in assuming that the limit of what LLMs can do exists on this (in image, 6.2 on the horizontal) dimension. Instead, we should look at the other axis, which in this example is called complexity. Complexity in this context refers to problems that require multiple steps, have vague problem statements, and require understanding when additional information is needed. Although it may seem obvious in hindsight, it is the problems with many small steps and decisions that are the real challenges for LLMs. However, with the development

[101] https://chatgpt.com/share/6752e8af-8114-8004-896b-88bab49e50ac The chatbot makes an incorrect interpretation by stating that § 2 also protects the functionality of program code. However, when it is requested without any prior example, it reaches the correct conclusion. https://chatgpt.com/share/6752eaf6-b64c-8004-b9e8-09098bbdbe37

of Chain of Thought (CoT) reasoning, it is difficult to predict how this will evolve in the future. These developments could significantly improve LLMs' capabilities in the legal domain.

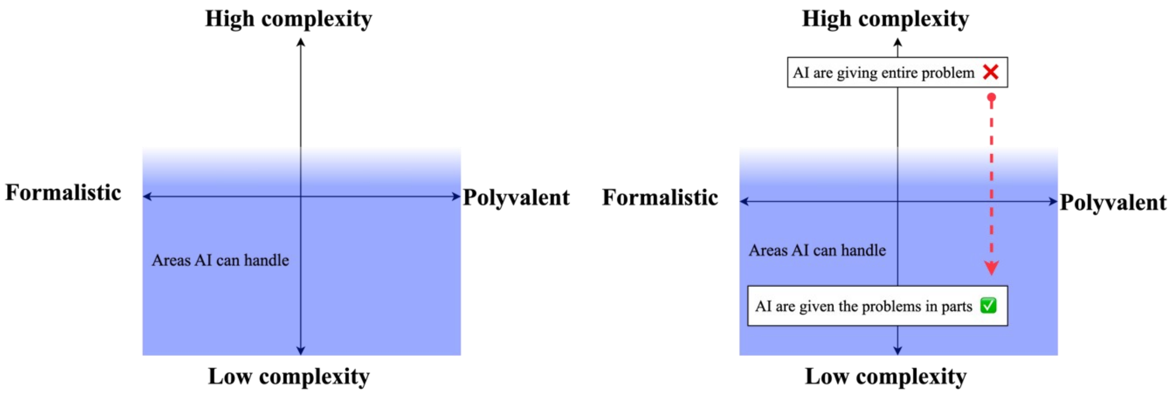


Figure 7.2 This figure provides an overview of all legal problems sorted along two dimensions. Areas within the blue region represent domains where LLMs have the capability to arrive at an appropriate solution. There is no clear boundary. Problems outside this region may become solvable through Chain of Thought (CoT) reasoning, which breaks down complex problems into smaller components that the AI model can solve independently while evaluating multiple possibilities before reaching a conclusion.

7.4. Key takeaways

I have demonstrated that the common criticism directed at AI, that it cannot understand abstract values, cannot evaluate values and rules against each other, and disregards principles and ethics, is not particularly strong criticism against LLMs. Instead, the transformer architecture in LLMs is well-adapted to handle these types of difficult questions. An LLM does not need to rely on a monocentric system where a single solution always exists, and everything can be broken down into components in a formalistic manner until everything can be solved logically.

Truly difficult centered problems are equally challenging for humans. With the stipulated theories as background, we do not need to find an AI system that outperforms humans in this regard, only one that mimics them. Although LLMs sometimes make incorrect judgments and applications, it is theoretically possible for them, using the same technology, to improve and reduce these problems.

However, I have demonstrated that there is another problem with LLMs applying law. It does not stem from centered complex problems. Instead, I show that LLMs encounter difficulties when faced with complex problems in the sense that they are extensive. If the legal problem contains many details and requires reasoning in multiple parts, the likelihood of the LLM making errors increases. Since LLMs only build on predicting the next word, these problems will increase progressively, and the final solution can differ significantly from a real lawyer's answer.

This can be resolved by improving LLMs' ability to apply the so-called Chain of Thought (CoT). This is currently where the development of AI stands today.

8. Real world test of LLM in legal reasoning

This research address question 2 regarding how well LLM models actually compare to aspiring lawyers.

To evaluate how well AI performs in the legal field, I conducted a test using current AI systems, which was evaluated by actual university lecturers. This was done to get a snapshot of AI's current capabilities in this domain. How well do today's AI systems perform in the legal field? It is clear that modern AI systems like ChatGPT, Claude, etc., cannot reproduce real-world court decisions in a way that matches human judges' output. A simple test demonstrates that these AI systems fall short of this standard. Furthermore, today's AI systems face several challenges, such as lacking access to all the information that influences a judge during a main hearing. This includes not only written submissions from both sides of a hearing but also various other elements like images, videos, and audio. Even factors such as how an attorney speaks, and moves can potentially influence a court.

This test was conducted on a smaller scale in simpler forms, partly because this is closer to the current capability level of modern AI systems, and partly due to practical considerations regarding how to provide AI systems, documents, and people for evaluation in the test.

8.1. The test

The specific purpose of this test was to determine if an LMM-based chatbot could perform the work of a legal professional. It became clear that AI does not yet have this capability, so a lower benchmark was chosen: to assess whether an AI system has progressed enough to perform at the level of a law student. Using Turing's method for determining if AI can think like humans, I created a test to see if AI could handle coursework similarly to students. Therefore, an exam from a course was selected where AI was tasked with creating an exam answer. Afterward, three university lecturers[102], who normally grade law program students, were asked to determine whether an answer was written by a student or by AI. If the lecturers cannot distinguish between AI and actual student answers at a rate significantly higher than chance, this would indicate that AI has reached a level where it can think like a law student. This in itself says less about quality. The purpose of the test was not to evaluate the quality of AI systems, but merely to determine if they can mimic humans. However, it was still interesting to include some indication of quality, hence such a question was included in the evaluation.

8.1.1. Material

The legal task that AI systems attempted was to write an answer to a law exam from the law program. Specifically, I selected the course HRO201 – Obligationsrätt (Contract Law), also known as Civilrätt 1 (Civil Law 1). This is a foundational course in law during the second semester of the law program in Gothenburg, which took place in spring 2024. There were

[102] Jakob Heidbrink, Associate lecturer in civil law at the department of law at Gothenburg univerity; Marianne Rödvei Aagaard, Associate lecturer in civil law at the department of law at Uppsala univerity and David Dryselius, Associate lecturer in law at the department of law at Lund university.

several reasons for choosing this course. It covers various legal methods, including direct, strict applications of laws, as well as analogies and partly ethical decisions. The difficulty level also appears to be sufficiently challenging for AI systems without being impossible.

This course has three regular exams taken by all students, with difficulty generally increasing with each exam. Students write these for over five hours without internet access and may only use course literature and a law book. The first two regular exams were used to test and evaluate chatbots, while the third regular exam was used for the actual test in this thesis.

As comparison material, I selected several law students' exam answers from the third regular exam. I chose answers from students who received the highest grade in the course.[103] Their names were never visible to the lecturers. Since one lecturer previously had examined some students in this class year, I was careful to ensure that he did not receive answers from students he had previously graded. This could otherwise have been a way to reveal that it was a student answer.

I also decided regarding spelling errors and minor grammatical mistakes. Since chatbots make significantly fewer such errors, this could potentially reveal their nature. A lecturer reading an exam answer with few spelling and reference errors might suspect it was created by AI. While this could be investigated further, I chose to correct all spelling and minor errors from the student answers to prevent them from being identified.

The number of minor errors could likely have been a factor in determining whether an answer was written by AI. However, that is not the question here. Generally, the more formal the legal text, the fewer spelling and grammatical errors occur.

8.1.2. The exam question[104]

The exam question presents a case study requiring an objective legal advisory opinion on conflicts and claims arising from a complex narrative. The case involves a newlywed couple, who entered into several contracts with different companies to arrange practical details for their honeymoon trip to the Seychelles.

The exam question is approximately 2 000 words long; A typical exam answer is about 3 500 words and is usually written within a 5-hour time limit. This is the final exam for the course and covers all learning objectives in the course.

This particular exam was chosen because it involves multiple legal relationships and presents an open-ended problem statement:

[103] I could have selected random students from all grade boundaries. However, this course contains 3 regular exams. By selecting only students who received an AB, I avoid choosing student responses that contain obvious errors. Generally, students with higher grades tend to be less stressed when writing their answers and therefore have more time to correct their linguistic errors.

[104] See entire exam question in Attachment 3: Exam question

”Take a reasoned position on how the conflicts and claims that you identify can or should be resolved legally. Assume an objective advisory role: you should not advocate for the couple, but rather inform them as objectively as possible about what solutions the law provides here. You should assume that all conflicts are resolved according to Swedish law.”[105]

First and foremost, the student must identify what constitutes a legal problem and what is legally relevant in this question. There is much information in the question that has no direct impact and must therefore be filtered out. Those who try to use traditional pre-LLM tools will not be able to find any help.[106] A Google search would only provide answers about the general application of rules but cannot take into account the specific circumstances of the question.

Even with simpler AI tools, it becomes difficult for students to use them effectively. The question is so extensive that earlier AI tools would not be able to handle it. Furthermore, the AI system needs access to both all available legal texts and relevant literature. Both were impossible to incorporate with earlier AI tools.

The question itself also encompasses legal issues of a more complex nature than those learned in basic law courses. It includes legal problems that lack explicit regulation and must therefore be solved through other means. The student must demonstrate the ability to use analogies and legal principles. Some parts of the question also involve reasonable assessments, where the student must determine what should be considered reasonable without any concrete answers provided by stipulated law.

Finally, this question is very lengthy and therefore requires the student to have the ability to create a clear response. This includes being able to organize their answer with clear headings and paragraphs and writing what the answer requires, neither too much nor too little.

8.1.3. The AI system in the test

Initial testing of the available AI system resulted in continuing with two chatbots: OpenAI’s *ChatGPT 4o*[107] and Antrophic’s Claude’s 3,5 *Sonnet*[108]. One of the biggest limitations was that the chatbot would have to be able to use all relevant regulations for the course (a PDF of 1000+ pages), and the relevant sections form the course literature. Though several chatbots have the ability to attach files to the prompts (probably using a mechanism known as RAG[109]), only ChatGPT 4o and Claude 3,5 Sonnet had de ability to upload the size required the option as well to know let the AI-system train on said material. Furthermore, several other chatbots had real

---

[105] Question from the exam. Translated by me. This question essentially captures exactly what is required for LJP.

[106] It is notably difficult to find pre-transformer systems that handle legal problems effectively. This may be a clear indication that these systems were never particularly successful. The ones I have managed to identify are *ROSS Intelligence*, *Ravel Law*, and *Lex Machina*. All of these systems appear to have been developed primarily to retrieve relevant regulations and case law, rather than to apply them to real-world circumstances.

[107] https://chatgpt.com/

[108] https://claude.ai/

[109] https://aws.amazon.com/what-is/retrieval-augmented-generation/

issues when trying to create an output similar to a student answer, instead of writing a short list.

To get reliable answers from chatbots, it became clear that I needed to break down the exam question into multiple steps, following a CoT approach. This was necessary for two reasons. First, modern AI is not capable of writing responses that match the length of a typical exam answer (which can be 7+ pages, while chatbots rarely provide more than 2 pages). Second, it is apparent that modern AI systems simply answer incorrectly when prompted with the question directly. They tend to respond too generally and barely apply the circumstances to the case at hand.

I solved this by creating a flowchart with different prompts that I ask step by step. The first prompt breaks down the question into several parts. Since almost all exam questions deal with multiple relationships, it was effective to divide the question accordingly. As an initial prompt, I asked the chatbot to identify which relationships were problematic. After that, I address one relationship at a time, asking questions that closely follow the RNTS model.[110] These steps, after trial and error, proved to be the best questions for guiding the AI model to write exam answers that most closely resemble those written by actual students.[111]

The sub-questions, however, could not be too specific, as could be seen as me answering the questions rather than the chatbot. The prompts were therefore specific enough to guide the chatbot in the right direction, but generally enough that the same questions could be used for all exam questions. This way, it is the model solving the questions, not me.[112]

One might criticize that I am assisting the model in solving the problem. To some extent, I am doing this by providing specific prompts. However, this is only temporary assistance. When asking the chatbot itself how it would solve the problem, it suggests similar prompts.[113] This shows that the AI model has the potential to perform this step independently. More recent models have also started developing ”reasoning" models, which largely involves giving itself an initial challenge to determine which steps to use, before executing these steps one by one.[114]

Just as actors make AI models available to the public, it is also very common for people to share useful prompts. This makes it relatively easy for someone to find prompts suitable for

---

[110] A common method used in the beginning of the law program. Stands for “Rättskälla”/”Legal Source”, “Norm”/”Rule or Principle”, “Tillämpning”/”Application” and “Slutsats”/”Conclusion”. Compare to the method described in Nääv and Zamboni, *Juridisk metodlära*, 30. Internationally similar methods are taught, such as “IRAC”.

[111] See Table 7.1 and Attachment 1: All evaluations

[112] Compare this to overfitting 6.4.1 Overfitting

[113] https://chatgpt.com/share/67373909-c7fc-8004-b862-45409e18db14 https://g.co/gemini/share/212dfe9233e4 See Figure 8.1.

[114] See ”Reasoning models,” 2023, accessed 5 december 2024, https://platform.openai.com/docs/guides/reasoning., ”Learning to Reason with LLMs,” 2024, accessed 10 October 2024, https://openai.com/index/learning-to-reason-with-llms/., ”Introducing Gemini: our largest and most capable AI model,” 2023, 7 december 2024, https://blog.google/technology/ai/google-gemini-ai/#performance, ”Meet Claude,” 2024, accessed 6 december 2024, https://www.anthropic.com/claude.

solving legal exams. This occurs without the person sharing the legal prompts having any idea which exams they will be used for.

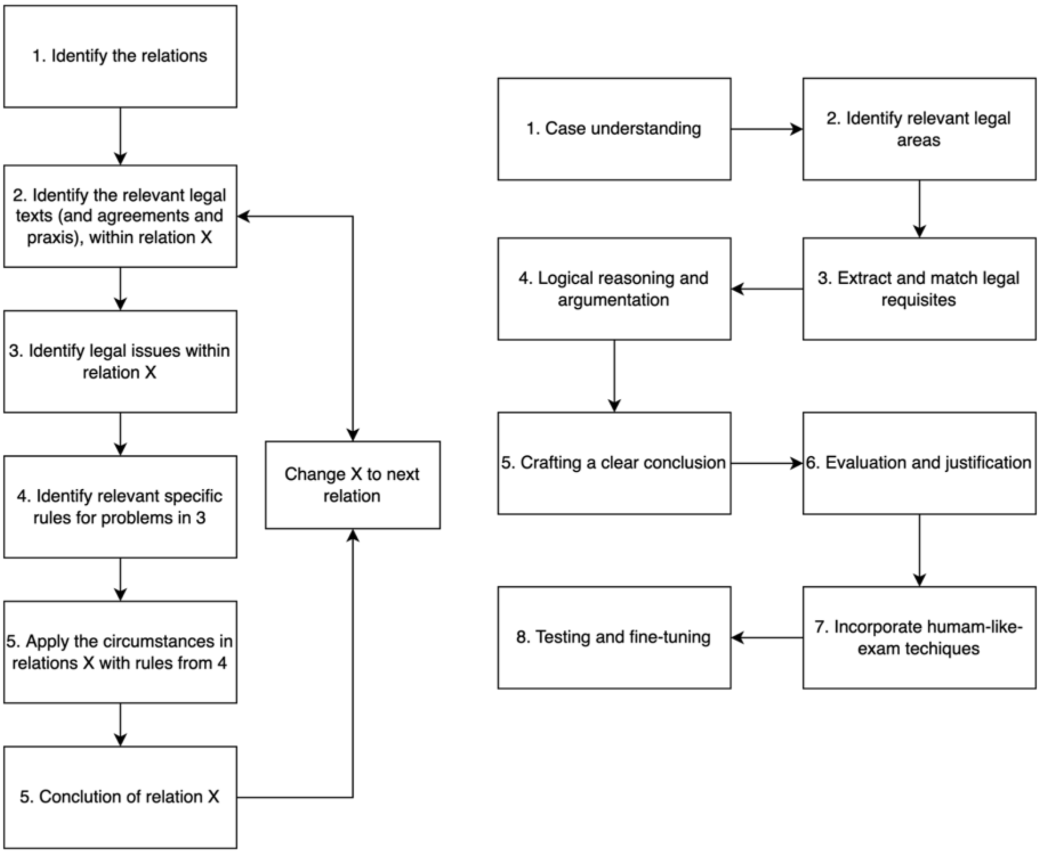


Figure 8.1: A flowchart I designed for solving exam questions (left) compared to a flowchart that ChatGPT generated when tasked with breaking down the problem (right).

8.2. Ethical concerns

This test includes not only exam answers written by chatbots, but also actual exam answers written by students at the University of Gothenburg. While these exam answers are publicly accessible upon request, they still constitute personal data.[115] My research aims to evaluate a Turing Test for AI-generated answers, a project that not only enhances my own learning as a student but also addresses a pressing public interest in assessing how effectively AI can mimic human student responses, especially given concerns that such technology might be misused for academic dishonesty. To ensure the protection of data subjects' rights, I have anonymized all data, stored it securely, and handled it in accordance with applicable data protection regulations. A comprehensive risk assessment has been conducted, demonstrating that the potential impact on individuals is minimal and that the benefits of this research outweigh the risks.

[115] General Data Protection Regulation (EU) 2016/679, Article 4.

There is a legitimate academic interest in accessing these materials for research purposes on behalf of the university.[116] No sensitive personal data[117] has been requested or accessed. The collection of personal data is within the policy of the University of Gothenburg.[118]

Personal data shall not be kept in a form that allows identification of data subjects for longer than necessary for the purposes for which the data is processed.[119] In the context of this study, exam answers, although publicly accessible. are considered personal data. Therefore, I have determined that this data will be retained only for the duration necessary for the research and subsequent validation of results. Following common practice in academic research at our institution, the data will be stored securely for a period of 10 years after the conclusion of the study, after which they will be permanently deleted or fully anonymized, ensuring that no individual can be re-identified.[120] For any questions or concerns regarding the handling of personal data, please contact me at gusbackgwi@student.gu.se[121].

Nevertheless, students might find it sensitive or feel violated that their exam answers have been used without their consent in this study. To address these concerns, all personal information (i.e., names, personal identification numbers, exam codes, and any other identifying information) has been completely redacted. Only the actual text of the exam answers remains. Furthermore, only answers from students who achieved the highest final grades were used in the test. Any potential spelling errors that could reveal a student's identity have been corrected before inclusion in the test. The only way to connect the exam answers to their authors would be through the content or writing style, which the appointed teachers for this test likely lack the capacity to distinguish. For the purposes of this study, I have determined that the research benefits outweigh any potential privacy concerns regarding the use of these student answers.

[116] Article 6(d).
[117] Article 9.
[118] https://studentportal.gu.se/dina-studier/rattigheter-och-skyldigheter/personuppgifter-inom-studierna?
https://studentportal.gu.se/dina-studier/rattigheter-och-skyldigheter/personuppgifter-inom-studierna?
[119] Article 5(1)(e),
[120] https://www.gu.se/pedagogik-specialpedagogik/var-forskning/datahantering
[121] If this mail has been terminated, please contact info@law.gu.se or dataskydd@gu.se.

8.3. The result

#	AI/Student	Model	%AI assessment
1	Student		33%
2	AI	ChattGPT 4o	33%
3	Student		0%
4	Student		0%
6	AI	Claude 3.5 Sonnet	67%
7	AI	ChattGPT 4o	67%
8	AI	ChattGPT 4o	0%
11	AI	Claude 3.5 Sonnet	0%
12	Student		33%
14	AI	ChattGPT 4o	67%
16	AI	Claude 3.5 Sonnet	33%
18	AI	Claude 3.5 Sonnet	33%
19	AI	ChattGPT 4o	100%
24	AI	ChattGPT 4o	67%

Table 8.1 A summary of lecturers guess of the answers. 100% in "%AI assessment" means all associate lectures assumed it was written by an AI.

Accuracy[122]	83%
Precision[123]	57%
Recall[124]	44%
F1[125]	56%

Table 8.2 A harmonic mean of accuracy, precision, recall and F1 between all associate lecturers.

A compilation of the results shows that it was not easy for the lecturers to determine which responses were AI-generated and which were written by real students. In this case, perhaps the most important parameter is *recall*, which shows how well the lecturers were able to identify AI responses. No lecturer managed to identify more than 56% of the responses correctly, and one lecturer had an identification rate as low as 11%. The combined average success rate for identifying responses was 49%, which is no better than random guessing.

It is worth noting that there were no precautionary measures in place when marking a response as AI in this case. If a response appeared 55% likely to be AI and 45% likely to be genuine, the lecturer probably chose to mark it as AI. In reality, however, teachers would require a higher degree of certainty to suspect that a response was written by a chatbot rather than a student, since this would essentially be an accusation of cheating. For this reason, the recall value, which

122 Total number of correctly classified responses divided by total number of responses.
123 Total number of correctly identified AI responses divided by total number of responses marked as AI by teachers.
124 Total number of AI responses correctly identified as AI divided by total number of actual AI responses.
125 The harmonic mean between precision and recall.

is the probability of teachers successfully identifying actual AI responses as AI, is likely much lower in reality.

Furthermore, the lecturers were aware that several responses would be AI-generated. This made them particularly attentive to distinctive characteristics that might indicate a response was written by a chatbot rather than a real student.

> *"I believe I recognize several turns of phrase typical of chatbots (such as the word "in conclusion" as an introduction to the concluding summary)."*[126]

> *"The lack of concrete application."*
> *"These types of non-answers are quite typical for AI bots."*
> *"Writes laws with capital letters and doesn't make referential errors."*[127]

The teachers also had the opportunity to grade all responses. This was a limited test as the lecturers could not dedicate equal time to each answer. The teachers spent an average of five minutes on each response. Additionally, they lacked a grading rubric to evaluate the answers against, instead providing a general grade for each response. The grades are therefore not entirely precise but can still provide valuable insights. A Turing test does not indicate anything about the quality of an answer. According to theories of American legal realism, there is no answer that is more "correct" than another. What judges rule becomes the correct interpretation, which is especially important for supreme courts. Regardless of whether one thinks a judge has "ruled incorrectly," their judgment becomes "the law." In this case, however, it was interesting to get an indication of how AI-generated answers compare. Students can, after all, perform quite poorly, and perhaps what is most interesting is whether modern AI systems can emulate students who receive passing grades.

Model	**Average Grade**	**All grades**	**Average grade**
ChattGPT 4o	17	B, AB, B	B+
Claude 3.5 Sonnet	10	Ba, B, U	B
ChattGPT 4o	3	U, B, U	U
ChattGPT 4o	13	B, B, Ba	B
Claude 3.5 Sonnet	10	Ba, U[128]	B
ChattGPT 4o	10	U, U, AB	B
Claude 3.5 Sonnet	17	Ba, B, Ba	B+
Claude 3.5 Sonnet	7	U, B, B	U
ChattGPT 4o	10	B, U, Ba	B
ChattGPT 4o	13	B, U, AB	B

Table 8.3 The collection of all grades given to all submissions written by AI, assessed by the lecturers.

[126] Jakob Heidbrink, commenting on why he suspects a response was written by AI. Translated by me.
[127] Comments from Marianne Aagard, Jakob Heidbrink and David Dryselius on another response. Translated by me.
[128] This test had only two evaluations.

In this test, teachers could grade responses on a scale from U–B–Ba–AB. These grades were converted to numerical values (0–10–20–30) to calculate averages. The grades showed significant variation, with some AI-written responses achieving AB quality while others failed. The average grade was 11, equivalent to a B, meaning AI generally pass the exams. This indicates that AI performs approximately at the same level as the average student, given that about 37% of students receive a B grade in the course and 17% fail.[129]

> *"Most AI-generated responses share a common flaw. They tend to describe rules and circumstances rather than apply legal principles (a problem also commonly seen among law students)."*
>
> *"Explains the rules but without the slightest indication of application.*
> *The answer indicates the area within which the solution can be found, but does not solve the problem."*[130]
>
> *"The answer is correct, but underdeveloped. There are missing explanations of how the assessment was carried out in several places and at multiple stages."*[131]

This presents a significant challenge for those attempting to use chatbots in legal contexts. Even with direct instructions and examples of how to apply legal principles, chatbots consistently avoid practical legal application.

A limitation of this test is that I was unable to conduct it using a fine-tuned chatbot. The training data for fine-tuning could have included high-scoring student responses, court decisions, or example texts demonstrating practical legal application. This would have particularly helped chatbots produce responses that better emulate applied legal reasoning. The training data that modern chatbots are currently trained on is likely more focused on descriptive law, such as law firms explaining rule interpretations, government agencies describing regulatory frameworks, and digital legal encyclopedias.

Although models that can be fine-tuned exist, they come with associated costs. The project lacked technical expertise, sufficient training data, and financial resources needed to test a fine-tuned version of an AI.

The varying quality of AI-generated responses supports the thesis that AI models struggle with extensive questions. In some cases, the AI model receives a good grade, while in others it performs significantly worse. This is partly due to the model selecting the wrong rule initially, which creates a cascading effect leading the AI model completely astray. However, in other instances, it arrives at more accurate conclusions.

---

[129] It should be noted that these statistics include students who have had two retake the exam. The number of students achieving a B grade after the first exam is likely significantly lower, considering that approximately 97 students, about 1/3 of the total students in the class, take the first retake exam.

[130] Comments from Marianne Aagard and Jakob Heidbrink on one response in regard to the grade. Translated by me.

[131] Comments from Jakob Heidbrink on one another response in regard to the grade. Translated by me.

Chatbots also tend to revert to purely descriptive explanations of laws, rather than demonstrating how these rules and principles should be applied. In very focused cases, one can guide the chatbot to show application, but when a question contains multiple sub-questions, there's a greater chance it will default to beginning with descriptions of legal provisions again. This is because these chatbots are primarily designed to function this way.

8.4. Conclusions from the test

The F1 score of 55% demonstrates that the lecturers, on average, had no ability to distinguish between student-written and AI-written answers, as this accuracy is equivalent to random guessing. The lecturers who participated in the test were prepared for the possibility of AI-written answers. This might differ in real-world scenarios where someone grading an exam may not be as vigilant about potential AI-generated responses. However, given the current widespread discussion about AI cheating, teachers should reasonably maintain heightened awareness for AI-written answers.

There was significant variation in the lecturers' assessment of answer quality. This could be attributed to the lecturers coming from different institutions, their slightly different approaches to grading exams, and the limited time they had to compare student answers against learning objectives. Despite these variations, the average score of AI-written answers reached a passing grade. To obtain a clearer understanding of the current capabilities of AI tools, more comprehensive and rigorous testing involving additional lecturers and teachers would be necessary.

In summary, this study suggests that AI may have reached a level where it can pass one of the most challenging courses in the Swedish law program. In any case, it is extremely difficult for lecturers to determine whether an answer was written by a student or a chatbot.

9. The future of LLM in law

As AI comes closer to emulating judges, the question of how to handle this becomes increasingly critical. If we accept that law is not impossible for LLMs to handle, we need to prepare accordingly. There are several potential areas within law that remain unexplored. For example, LLMs that are fine-tuned with specific Swedish legal texts could suddenly perform significantly better than today's non-fine-tuned models.

Law is also a field that varies significantly from country to country. Besides having completely different regulations, countries have their own legal systems and cultural differences. This contrasts with mathematics, physics, chemistry, biology, psychology, and other advanced subjects that are much more uniform across countries. This means we might see a delay before AI becomes proficient in handling Swedish law. This differs from the situation in the US, where most modern AI systems are developed and where there might be more investors in the AI field. This could mean AI will more quickly become adept at handling American law. In the EU, with its stricter data protection laws and jurisdictions that differ more than American states, it may take longer before a good AI model is developed.

There is also a possibility that AI system development will reach a plateau. The development that has occurred between 2017 and now has been explosive, but this might have been temporary. New limitations might come into play, making AI systems increasingly difficult to develop at the same pace. This could be due to pure technical limitations, such as exponentially increasing data capacity requirements becoming harder and harder to satisfy. It could also be due to other surrounding factors. Increasing personal data protection measures might place limitations on developing these systems. Training data used for these systems is receiving stronger protection and recognition against use in ML. Pure legal compliance regulations are making it increasingly costly to develop larger and more advanced AI systems.

10. Conclusion

The following is a summary of the conclusion to questions presented in 2. Problem and purpose. It is divided into question 1 in section 10.1 Theoretical challenges with LLM used in law and question 2 in section 10.2 The capabilities of LLM today. Finally, a conclusion will be presented drawing from both of these findings in 10.3 LLM in the field of law.

10.1. Theoretical challenges with LLM used in law

Are there any challenges within Legal Problem Solving that LLM-based AI models cannot or find very difficult to solve?

In my analysis, I concluded that while LLM-based AI models demonstrate impressive language skills and the ability to identify and apply legal rules, they face significant challenges in legal problem solving. One key issue is that legal reasoning often requires breaking down a problem into a series of logically connected steps. Since LLMs generate text by predicting the next word based solely on statistical likelihoods, they lack an inherent mechanism to revisit and correct earlier steps. This means that an initial misinterpretation can propagate errors throughout the entire reasoning process.

From the perspective of American Legal Realism, where the law is essentially what jurists determine it to be, AI models largely function by mimicking jurists. Previous research has highlighted difficulties for AI systems because the legal system is far more complex than simple "if-then" statements. Polyvalent theories demonstrate how values from multiple parts of society play important roles. This illustrates the complexity of the legal system, where jurists must constantly consider multiple rules simultaneously and weigh them against each other. Often, there exists more than one "correct answer."

This posed significant challenges for AI systems before transformers. However, this thesis, along with theoretical analyses and certain examples, has demonstrated that this is less problematic for LLMs. The transformer mechanism simply allows AI systems to much more closely resemble human jurists in handling legal matters in a complex world. The weights in neural networks and their dimensions can emulate complex relationships, including ethics, trade-offs, and reasonableness assessments.

The research has shown that LLMs build their reasoning on intuition, similar to the hunch theory. This enables AI systems to handle problems that are difficult to solve concretely but rather require intuition. However, this also reveals the first problem with LLMs: they struggle to correct themselves and to handle questions of high complexity.

It is not individual, focused questions that pose problems for LLMs, but rather comprehensive questions with many potential sub-questions and facts that make it difficult for the LLM to know where to begin, which issues are important, and how to correct itself. This is why these types of questions are more challenging for LLM models than focused difficult questions, such as ethical reasonableness assessments, unclear rule applications, analogies, etc. While LLM-based AI can handle many aspects of legal problem solving, they are particularly challenged by problems that require a high degree of sequential reasoning, nuanced interpretation, and contextual adaptation. Future improvements, especially in enhancing their chain-of-thought capabilities, may help to overcome some of these limitations, but as it stands, these challenges are significant obstacles to achieving full legal expertise with current AI technology.

10.2. The capabilities of LLM today

Are LLM-based AI models capable of handling a basic course in the Swedish law program?

The test conducted in this thesis has indicated serious conclusions. A modern, free-market AI chatbot may be capable of passing one of the hardest courses at one of the Swedish law schools without the lecturer being able to detect it. The Turing test used in this case demonstrates that modern chatbots have the ability to imitate students. The quality of the answers is not higher than passing grade. Further investigation is necessary to definitively determine whether AI models can pass the course. Nevertheless, this provides a good indication.

A student who wants to cheat using modern AI has the possibility to pass exams today with minimal risk of detection. This is based on either the exams being take-home assignments, or the security measures for on-site exams are not sufficiently robust, allowing students to still gain access to chatbots (there are examples of students who have managed to do this).[132] Institutions that rely on their ability to detect AI-written responses should reconsider their approach, as this study has shown that teachers on average cannot distinguish them better than random chance.

Interestingly, even though I explicitly asked AI to answer an exam question and even tried to make it look like it was written by a student, not once did any AI bot stop me and inform me

[132] Dahlström, *Student använde AI på tenta – frias från anklagelser om fusk*.

that what I was doing was unethical. This is otherwise something that chatbots are apparently programmed to do in other unethical cases.[133]

The test also indicates that AI is approaching the ability to solve legal problems without being trained on the exact circumstances in question. These tests demonstrate, although not completely conclusively in this particular case, that modern LLMs based on the transformer model have the potential to assist with legal matters.

10.3. LLM in the field of law

Those who study jurisprudence and computer science understand that it would likely require something like an LLM with transformers to even begin to master legal reasoning. Law is too full of exceptions, dependent on reality, and constantly changing to be mastered by purely symbolic AI.

There are significant concerns about AI development. Many discussions focus on the dangers of AGI and what is called singularity, when AI becomes smarter than humans and could potentially create better versions of itself, leading to potentially devastating circumstances. However, many also see more immediate threats. A common concern is AI taking over occupations, something some economists now predict.[134] This discussion has repeated itself with every major technological advancement. While this thesis about this aspect, we need to avoid having a one-sided view.

AI development is particularly interesting in the legal field because it is an area that has seen relatively little change in professional practice despite all major innovations throughout history. While most lawyers today use computers and electronic references, court proceedings essentially function the same way they have for the past hundred years. Just as someone with a legal problem would consult an expert, a lawyer, a hundred years ago, they need to do the same today. Those who seek answers online might find some solutions and advice, but these can often be misleading and even lead to greater legal harm than if they had consulted a lawyer. One reason why the internet has not been able to help those needing legal advice is that there are too many varying circumstances that need to be properly applied to legal problems.

Furthermore, there are no programs or scripts that can significantly assist lawyers. While many professions have seen major productivity boosts through assistive tools, creating similar systems for lawyers has proven difficult.[135] Only more efficient databases, such as online regulations and simple calculation models for things like property division and inheritance, have proven useful.

[133] https://chatgpt.com/share/6752d011-90f4-8004-aa8f-d1e2ecd71f8a https://g.co/gemini/share/dc0ab311ab7a https://chatgpt.com/share/6752d088-ad70-8004-a68b-a32a853f4f9b https://gemini.google.com/share/f0ca7e464108
[134] *People at Work 2024: A Global Workforce View*, 41.
[135] Chui, Hazan, Roberts, Singla, Smaje, Sukharevsky, Yee and Zemme, *The economic potential of generative AI*, 45.

Therefore, the legal field stands at an interesting crossroad, as it appears to be truly challenged by new technology for the first time. Lawyers who once held exclusive expertise are now facing a challenge. Individuals needing help with legal problems might potentially get satisfactory assistance through AI models.

If this means that people who actually have rights can better stand up for themselves and obtain justice, perhaps we should welcome this development. AI can thus be seen as a technology that truly helps address the eternal problem of "access to justice." You no longer need to pay enormous sums to claim the rights you are entitled to.

AI models can also help more people understand "applicable law." This means that individuals, without having to take their problems to court and eventually the supreme court, can get more accurate predictions of how a legal case "would be judged." When American legal realists talk about law being only what judges decide, one can criticize this for missing all those who understand a rule that judges have not ruled on. AI that becomes skilled at predicting supreme court decisions will help reduce this discrepancy. Individuals who struggle to determine what to do in specific situations can form an understanding through AI and thus avoid costly court processes to find out the same information.

The legislative branch can also benefit from AI that becomes better at predicting potential legal cases. The legislative power could test potential law changes using AI to better understand how they will be used once enacted. Through this process, the legislative power can better catch potential loopholes and get a clearer picture of how the law works in practice. Written law and law in practice thus become somewhat more intertwined.

There are, of course, negative aspects as well. Those who want to exploit the law might also possibly find loopholes. What is perceived as criminal but, due to technicalities, is not illegal might become more visible to those individuals.

Most experts argue that AI systems lack genuine understanding of reality and can only comprehend the world as we teach it to them.[136] An AI system that fills gaps in one legal domain by drawing analogies from another may unknowingly make inappropriate comparisons between fundamentally different areas. This can occur when there are superficial similarities between legal domains that are actually quite distinct in nature. A clear example of AI misinterpreting reality due to misleading similarities is the Uber self-driving Volvo accident. Elaine Herzberg was walking her bicycle across a road when the autonomous vehicle incorrectly classified her as a cyclist, leading to erroneous trajectory calculations. This miscalculation resulted in the first fatal traffic accident caused by a self-driving car.[137]

---

[136] Gärdenfors, *Kan AI tänka? : om människor, djur och robotar*, 108.

[137] The exact reasons for the crash seem to be conflicting. Some sources state it mistook the pedestrian for a plastic bag, while others claim that the sensors detected her as a bicycle very late, see ibid 203. *Preliminary Report Highway HWY18MH010*.

Moreover, AI systems do not emerge on their own. They require many people to help develop these systems. Data developers gain significant power as the last people to work on AI before they begin interacting with the public. If AI systems are to handle legal problems, it is probably a good idea to have more lawyers involved who can ensure that, for example, the right training data has been used in the learning process. Just as in my test, lawyers need to be involved in the evaluation process. Without lawyers commenting on AI handling legal matters, we have no idea how well these systems actually work.

As the ELIZA example demonstrates, it is not sufficient to simply have skilled data professionals collect legal materials.[138] Through this process, one might create AI that appears to explain legal concepts well, but in reality, produces no substantive legal content. An interesting future experiment would be to have non-legally trained individuals perform the same test to see if they can determine which responses were written by whom. Such a test might well reveal that non-legally trained individuals find AI responses more compelling, as chat models are designed to provide satisfying answers for the general public rather than legally sound responses.

This demonstrates that even as AI approaches mastery of the *art of law*, there remains a strong societal interest in including legal professionals in this development. In retrospect, we might view today's AI developments similarly to Y2K, but just as with Y2K, our concerns may actually contribute to implementing precautionary measures that prevent any catastrophic outcomes.

Throughout this thesis, I may appear to present a very positive view of AI in law. Positivity is a relative concept and given the predominantly negative perception of AI's development in the legal world, this perspective might seem positive. However, my main point has not been to demonstrate that AI can do everything – this is especially true with today's AI systems, which are far from outperforming the most skilled lawyers. Instead, my point has been to show that we need to be more open to the possibility that AI might master legal work. We need to be more receptive to the idea that humans might not be the only ones capable of mastering law.

Previously, we believed that chess would never be mastered by machines. Chess was seen as the ultimate challenge demonstrating human intelligence. It was not just about calculating the best move; one had to play a psychological game against their opponent and understand their thinking. This belief ended abruptly when Deep Blue defeated grandmaster Kasparov in 1997. A similar scenario occurred decades later when Google mastered the game of GO in 2016, a game previously thought impossible to master with static AI models.

Similarly, lawyers need not be closed off to the idea that certain tasks can only be handled by humans. The technology of transformers has shown that we need to rethink how we view AI. It is no longer just a logical machine calculating everything in minute detail to arrive at the

[138] See 6.2.1 ELIZA

most probable outcome. Instead, we must also see AI as these chaotic systems that can recognize patterns without us understanding how these connections fundamentally work.

If we maintain a more open and receptive attitude to the possibility that the sacred domain of law will one day be challenged by AI, we can better prepare ourselves. We can be more open to shaping legal AI to follow the kind of law we want to exist. We can ensure that it behaves ethically, remains accessible to more people, and avoids biases.

References and bibliography

Domstolsstatistik 2023, 2023.

People at Work 2024: A Global Workforce View, The ADP Research Institute 2024.

Preliminary Report Highway HWY18MH010, National Transportation Saftey Board.

Abbott, R., The reasonable robot : artificial intelligence and the law, Cambridge : Cambridge University Press 2020.

Agency, U. C. I., The Enigma of Alan Turing, 2015, accessed accessed 10 october, 2024, https://www.cia.gov/stories/story/the-enigma-of-alan-turing/.

Aletras, N., D. Tsarapatsanis, D. Preotiuc-Pietro and V. Lampos, Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective, San Diego: PeerJ. Ltd 2016, e93.

Bench-Capon, T., Argument in Artificial Intelligence and Law, 1997, 249-261.

Bengtsson, J., Göteborgs universitet läxar upp student – stängs av efter misstänkt AI-fusk Göteborgs-Posten, 2024, accessed, 20 November, 2024, https://www.gp.se/nyheter/goteborg/goteborgs-universitet-laxar-upp-student-stangs-av-efter-misstankt-ai-fusk-.32933d07-be22-4e38-8b08-ed95f8a95272.

Bird, J., Engineering Mathematics, Newnes 2007.

Bladini, M., Om emotioners vara eller icke vara — ett exempel från processrätten, 2019.

Bojic, L., P. Kovacevic and M. Cabarkapa, GPT-4 Surpassing Human Performance in Linguistic Pragmatics, Ithaca: Cornell University Library, arXiv.org 2023.

Chen, Y., Q. Fu, Y. Yuan, Z. Wen, G. Fan, D. Liu, D. Zhang, Z. Li and Y. Xiao, Hallucination Detection: Robustly Discerning Reliable Answers in Large Language Models, Ithaca: Cornell University Library, arXiv.org 2024.

Chui, M., E. Hazan, R. Roberts, A. Singla, K. Smaje, A. Sukharevsky, L. Yee and R. Zemme, The economic potential of generative AI, 2023.

Claude, Meet Claude, 2024, accessed accessed 6 december 2024, https://www.anthropic.com/claude.

Dahlberg, P., Myten att en polisanmälan om övergrepp är förtal drabbar kvinnor, 2021, accessed accessed 23 November 2024, https://arbetet.se/2021/03/17/myten-att-en-polisanmalan-om-overgrepp-ar-fortal-drabbar-kvinnor/.

Dahlström, E., Student använde AI på tenta – frias från anklagelser om fusk Göteborgs-Posten, 2024, accessed accessed 21 November, 2024, https://www.gp.se/nyheter/goteborg/student-anvande-ai-pa-tenta-frias-fran-anklagelser-om-fusk.47b3b9bb-96a2-4b6e-8131-86fd2be1fa50.

European Commission, AI Act enters into force, 2024, accessed accessed 29 november, 2024, https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en.

Foster, D., Generative Deep Learning, Sebastopol: O'Reilly Media, Incorporated 2023.

Google, Google reCAPCHA, Creation of Value, 2024, accessed accessed 2 December, 2024, https://www.google.com/recaptcha/intro/?zbcode=inc5000#creation-of-value.

Gustafsson, H., Rättens polyvalens: en rättsvetenskaplig studie av sociala rättigheter och rättssäkerhet, Diss. Göteborg : Univ., 2003 2002.

Gärdenfors, P., Kan AI tänka? : om människor, djur och robotar, Stockholm : Fri tanke 2024.

Häggström, O., Tänkande maskiner, Stockholm : Fri tanke 2023.

IBM, Deep Blue, accessed accessed 11 nov, 2024, https://www.ibm.com/history/deep-blue.

Kahneman, D. and G. Klein, Conditions for Intuitive Expertise: A Failure to Disagree, WASHINGTON: American Psychological Association 2009, 515-526.

Kang, X., L. Qu, S. Lay-Ki, A. Trakic, Z. Terry Yue, P. C. Emerton and G. Grant, Can ChatGPT Perform Reasoning Using the IRAC Method in Analyzing Legal Scenarios Like a Lawyer?, Ithaca: Cornell University Library, arXiv.org 2023.

Kelsen, H. and M. Jestaedt, Reine Rechtslehre : Einleitung in die rechtswissenschaftliche Problematik (Studienausgabe der 1. Auflage 1934), Mohr Siebeck 2020.

Kleinberg, J., S. Mullainathan and M. Raghavan, Inherent Trade-Offs in the Fair Determination of Risk Scores, Ithaca: Cornell University Library, arXiv.org 2016.

Kronblad, C., J. E. Pregmark and R. Berggren, Difficulties to digitalize: ambidexterity challenges in law firms, 2023, 217-236.

Llewellyn, K. N., The Bramble Bush: on our Law and its Study, New York (State): Oceana Publications, 1951 1951.

Marquis, P., O. Papini, H. Prade, H. Prade, P. Marquis and O. Papini, A Guided Tour of Artificial Intelligence Research: Volume III: Interfaces and Applications of Artificial Intelligence, Cham: Springer International Publishing AG 2020.

Michanek, G., Den svenska miljörätten, Uppsala : Iustus 2017.

Mitchell, M., Artificial intelligence : a guide for thinking humans, London : Pelican an imprint of Penguin Books 2019.

Mnih, V., K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller, Playing Atari with Deep Reinforcement Learning, Ithaca: Cornell University Library, arXiv.org 2013.

Modak-Truran, M. C., A Pragmatic Justification of the Judicial Hunch, 2001.

Nelson Elhage, T. H., Catherine Olsson, Nicholas Schiefer, Tom Henighan, Shauna Kravec, Zac Hatfield-Dodds, Robert Lasenby, Dawn Drain, Carol Chen, Roger Grosse, Sam McCandlish, Jared Kaplan, Dario Amodei, Martin Wattenberg∗,Christopher Olah, Toy Models of Superposition, 2022, accessed accessed 29 december 2024, https://transformer-circuits.pub/2022/toy_model/index.html.

Nelson, W. E., Brown v. Board of Education and the Jurisprudense of Legal Realism, 2015.

Nielsen, M. A., Neural Networks and Deep Learning, Determination Press 2015.

Noll, G., AI, digitalisering och rätten : en lärobok, Lund : Studentlitteratur 2021.

Nääv, M. and M. Zamboni, Juridisk metodlära, Lund : Studentlitteratur 2018.

Oliver Wendell Holmes, J., The Path Of The Law, Simon Brown 1897.

OpenAI, Learning to Reason with LLMs, 2024, accessed accessed 10 October 2024, https://openai.com/index/learning-to-reason-with-llms/.

OpenAI, Reasoning models, 2023, accessed accessed 5 december 2024, https://platform.openai.com/docs/guides/reasoning.

Pasquale, F., New Laws of Robotics, Harvard University Press 2020.

Prakken, H. and G. Sartor, Law and logic: A review from an argumentation perspective, AMSTERDAM: Elsevier B.V 2015, 214-245.

Richards, D., When Judges Have a Hunch: Intuition and Experience in Judicial Decision-Making, Stuttgart: Franz Steiner Verlag 2016, 245-260.

Rissland, E. L., Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning, NEW HAVEN: The Yale Law Journal Company 1990, 1957-1981.

Rogers, S., A first course in machine learning, Boca Raton : CRC Press 2016.

Rotolo, A. and G. Sartor, Argumentation and explanation in the law, Frontiers Media S.A 2023, 1130559-1130559.

S, B., S. Kadry, R. K. D and P. Aruchamy, Generative AI and LLMs: Natural Language Processing and Generative Adversarial Networks, 2024.

Sartor, G., Doing justice to rights and values: teleological reasoning and proportionality, Dordrecht: Springer Netherlands 2010, 175-215.

Scherer, M., Artificial Intelligence and Legal Decision-Making: The Wide Open?, Alphen aan den Rijn, The Netherlands: Kluwer Law International 2019, 539-573.

Schwab, K., The Fourth Industrial Revolution: what it means, how to respond World Economiv Forum, 2016, accessed accessed 15 November 2024, https://www.weforum.org/stories/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/.

Solberg Søilen, K., A Review of the Knowledge Worker as Prompt Engineer: How Good is AI at Societal Analysis and Future Predictions?, 2024, 6.

Strömholm, S., Rätt, rättskällor och rättstillämpning : en lärobok i allmän rättslära, Stockholm : Norstedts Juridik 2020.

Sundar Pichai, D. H., Introducing Gemini: our largest and most capable AI model, 2023, accessed, 7 december 2024, https://blog.google/technology/ai/google-gemini-ai/#performance.

Tealab, A., Time series forecasting using artificial neural networks methodologies: A systematic review, Elsevier B.V 2018, 334-340.

Teubner, G., Law as an Autopoietic System, Blackwell Publishers 1993.

Towes, R., Transformers Revolutionized AI. What Will Replace Them? Forbes, 2023, accessed accessed 25 november 2024, https://www.forbes.com/sites/robtoews/2023/09/03/transformers-revolutionized-ai-what-will-replace-them/.

Turing, A., Computing Machinery and Intelligence, 1950.

Turner, J., Robot rules : regulating artificial intelligence, Cham, Switzerland : Palgrave Macmillan 2019.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, Attention Is All You Need, Ithaca: Cornell University Library, arXiv.org 2023.

Wacks, R., Understanding jurisprudence : an introduction to legal theory, Oxford : Oxford University Press 2017.

Wang, Z. J., R. Turko, O. Shaikh, H. Park, N. Das, F. Hohman, M. Kahng and D. H. Polo Chau, CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization, LOS ALAMITOS: IEEE 2021, 1396-1406.

Weizenbaum, J., ELIZA: a computer program for the study of natural language communication between man and machine, New York, NY: Association for Computing Machinery 1983, 23-28.

Wittgenstein, L., Filosofiska undersökningar, Stockholm : Thales 2012.

Xie, C., L. Xunying, W. Yu, A. Ragni, J. H. M. Wong and M. J. F. Gales, Exploiting Future Word Contexts in Neural Network Language Models for Speech Recognition, PISCATAWAY: IEEE 2019, 1444-1454.

Attachment 1: All evaluations

Student/AI	Model used	Average score	All grades	Average grade	%AI
Student		20	B, Ba, AB	Ba	33%
AI	ChattGPT 4o	17	B, AB, B	B+	33%
Student		27	AB, AB, Ba	Ba+	0%
Student		30	AB, AB, AB	AB	0%
AI	Claude 3.5 Sonnet	10	Ba, B, U	B	67%
AI	ChattGPT 4o	3	U, B, U	U	67%
AI	ChattGPT 4o	13	B, B, Ba	B	0%
AI	Claude 3.5 Sonnet	10	Ba, U	B	0%
Student		27	Ba, AB, AB	Ba+	33%
AI	ChattGPT 4o	10	U, U, AB	B	67%
AI	Claude 3.5 Sonnet	17	Ba, B, Ba	B+	33%
AI	Claude 3.5 Sonnet	7	U, B, B	U	33%
AI	ChattGPT 4o	10	B, U, Ba	B	100%
AI	ChattGPT 4o	13	B, U, AB	B	67%

Attachment 2: Prompts used in the test

1. Identify relations

Utifrån uppgiften i frågan - Identifiera och redogör för relationerna mellan parterna i frågan samt deras juridiska relevans.

Skriv i denna form:

"Köpet av X mellan privatpersonen A och företaget B"

"Avtalet mellan personen E och personen F"

2. Identify relevant regulations

Vilka av rättsliga regler är relevanta och tillämpliga gällande RELATION? Ta jämför lagrummens tillämplighetsparagrafer mot omständigheterna i detta fall. Gör en bedömning i detta fall. Exempelvis:

”I detta fall är köplagen inte tillämplig eftersom det handlar om ett köp mellan två näringsidkare av en tjänst (KöpL 1 §). Konsumenttjänstlagen är heller inte tillämplig eftersom den endast är tillämplig när eftersom båda avtalsparterna är näringsidkare (KTjL 1§). I stället gäller allmänna avtalsrättsliga principer samt Avtalslagen för de aktuella avtalsfrågorna (se *Avtalsrätt I*, Adlercreutz, Gorton & Lindell-Frantz, s. 35).”

”I detta fall är lagen om handelsagentur tillämplig eftersom Anna, som agent, arbetar självständigt och varaktigt för att sälja företagets varor till kunder för huvudmannens (företagets) räkning (1 § HAL)"

3. Identify legal issues

Identifiera de rättsliga problem och centrala frågeställningar som är relevanta gällande RELATION, och redogör för de juridiska frågor som behöver besvaras för att förstå parternas rättigheter och skyldigheter. Beskriv vad parterna eftersträvar och vilka intressen de vill skydda eller uppnå inom avtalsrelationen. Om det finns delfrågor som måste besvaras innan en slutlig lösning kan nås, ta upp dessa och deras relevans i relation till parternas mål. Håll dig kort och koncist.

4. Identify relevant specific regulations

För varje identifierat juridiskt problem ovan, redogör för vilka specifika lagrum och rättsliga principer som kan vara tillämpliga. Ange exakta paragrafhänvisningar, utan att tolka eller lösa det rättsligt. Om ett problem saknar ett direkt tillämpligt lagrum för ett specifikt problem inom det aktuella rättsområdet, påpeka detta tydligt utan att försöka hantera det rättsligt.

5. Apply circumstances

Tilllämpa de tillämpliga lagrum från ovan på omständigheterna i alla problemområden som nämnt innan kopplat till RELATION. Utred delfrågorna först.

Gör bedömningar, exempelvis avvägningar och kom till en slutsats.

För problem som saknar direkt tillämplig lag, använd juridiska metoder för att lösa det på ett så bra sätt som möjligt.

6. Fill argumentation with sources

Sök igenom din kunskap. Finns det något i den som kan styrka eller fylla ut din påstående i senaste meddelande? Vid hänvisning till litteratur i din kunskap, skriv alltid ut sidanvisning. Se över så alla påståenden har en källa.

7. Conclusion

Skriv nu ett väldigt utförligt slutgiltigt svar på denna fråga gällande RELATION, som om du svarade på en tenta. Använd dig av enkla rubriker så svaret blir lätta att följa. Skriv i stycken, använd inte punktlistor eller liknande struktur.

Attachment 3: Exam question

Inför bröllopsresan

Bröllopet är efter alla vedermödor och allt som gått snett (faktiskt) äntligen över. Sam och Mino är lyckliga att ha bytt ringar, även om de kan önska att allt skulle ha varit lite mindre komplicerat och lite mindre stökigt. Nu stundas bröllopsresa! Det nygifta paret ska tillbringa två veckor på Seychellerna – och inga uppladdningar på sociala medier om något de gör gemensamt under tiden, har Sam lovat!

Innan de två åker iväg, passar Sam därför under tiden mellan det privata bröllopet och den privata bröllopsresan på att fylla sina kanaler med innehåll och att förbereda fansen bland följarna på att det nu kommer att gå lite saktare i flödet. Hon lägger alltså upp filmer och bilder i en rasande takt: bland annat går hon igenom den stora skörden av bröllopspresenter och presenterar dem för följarna.

Det är flera företag och även följare, förutom förstås familjen, som passat på att skicka presenter.
Sam visar upp det mesta. Bland annat visar hon en uppsättning smycken – ett halsband med matchande örhängen till ett värde om 50 000 kronor – som hon fått av en juvelerarkedja med vilken hon samarbetat ett tag och en äkta tavla som en följare som i konstkretsar anses vara ”up and coming” målat särskilt för henne och Mino. Sammantaget är det presenter om ett värde på – uppskattar Sam – 200 000 kronor som Sam visar upp.

Allt ordnas

Inför själva resan ordnar paret med det praktiska. Mino bokar i sitt eget namn en motorbåt till vilken paret ska ha tillgång under semestern. Båten bokas hos ett svenskt företag som skyltar med att det levererar ”kundkvalitet och service på Sveriges högsta nivå”. Mino lägger i samband med att avtal sluts in ett betalningsuppdrag hos parets bank: hyran utgör ett fast pris om 50 000 kronor och ska betalas en månad efter bokningen, vilket infaller en vecka efter det att paret är hemkommen från semestern.

Sam å sin sida sluter, i sitt bolags namn (hon är VD i sitt aktiebolag, vilket enligt bolagsordningen ska syssla med ”medieverksamhet”), avtal med ett magasineringsföretag. Avtalet går ut på att paret ska få obehindrad tillgång till utrymmen i företagets fastighet där det kan ställa in sådant som det inte vill ha i bostaden. Att Sam sluter avtalet i bolagets namn har mest med det praktiska i stunden att göra: Sam tänker sedan dokumentera hur allt gått till när bolaget tog hand om hennes och Minos egendom, och att av parets privata pengar återbetala de pengar som bolaget fått betala till magasineringsföretaget.

Givet att estetik och inredning är viktiga för både Sam och Mino, är det ganska mycket som åtminstone tills vidare inte ska få plats i den gemensamma bostaden: de två vill ha tid att gemensamt fundera på var i bostaden som vad ska få stå eller hänga, och också vad av alla presenterna som kanske är så ”fult” att det diskret ska säljas efter hemkomsten från bröllopsresan. Bland annat ställs tavlan som Sams följare har målat i utrymmet hos magasineringsföretaget.

Mino sluter dessutom ett bevakningsavtal med ett säkerhetsföretag. Enligt avtalet – som anger att uppdraget gäller en privatbostad – ska företaget en gång om dagen checka att inget hänt med lägenheten där paret bor. Dessutom installeras ett larm. Enligt avtalet ska företagets personal, om och när ett larm utlöses, inom tio minuter vara på plats i lägenheten och kontrollera vad som hänt. Dessutom ingår branddetektion i avtalet.

Ett nytt beslut

När allt är fixat och väskorna packade ska paret iväg tidigt en måndag morgon. På söndagen dessförinnan pratar Sam och Mino om den kommande semestern och om hur skönt det ska bli att efter alla problem och konstigheter med bröllopet och dess efterdyningar komma i väg och slappa i två härliga veckor. Samtalet utvecklas sådant att paret kommer fram till att det egentligen inte har någon större lust att vara så där superaktivt på sin semester. Visst, Sam måste lägga ut
åtminstone några filmer och häftiga bilder – ”en influenser är aldrig ledig”, som hon brukar säga – men i huvudsak vill både Sam och Mino samla krafter inför den långa och tråkiga svenska hösten. I synnerhet det där med båten låter inte alls lika lockande längre: hotellet där paret ska bo tillhandahåller taxibåtar och även lånebåtar, förvisso inte i den lyxklass som den båt som Mino bokat, men hotellets båtar duger helt klart för att ta sig runt och också för att göra utflykter. De två bestämmer sig för att avboka båten. Mino fixar ett meddelande till uthyrningsföretaget vid kl. 22 på kvällen; sedan är det dags att lägga sig och vakna tidigt för att hinna med flyget. När paret redan sitter på flygplanet anländer till Minos mejladress ett meddelande om att bolaget noterat och accepterat avbeställningen, men att det på grund av att avbeställningen var så sen enligt avtalets villkor 21 debiterar en avbokningsavgift på tio procent av priset eller 5 000 kronor.

En ände med förskräckelse

Semestern blir precis som tänkt och riktigt härlig. Sam och Mino återkommer efter två veckor, utvilade och glada över det lyckade avslutet på en sorgligt komplicerad bröllopssaga.

Men det visar sig vara för tidigt för att ropa hej.

När paret kommer hem, visar det sig att de haft inbrott i lägenheten. Tjuvarna har varit skickliga och lyckats forcera dörren med minimal skada. De stängde också dörren efter sig, så att det utifrån inte alls är uppenbart att något hänt: Sam och Mino reagerar först när de skjuter in nyckeln i låset och nyckeln fastnar. När de två väl tagit sig igenom den skadade dörren och in i lägenheten, möts de av totalt kaos. Några konstkeramikföremål ligger i skärvor på golvet och

Sams smycken som hon fått till bröllopet är borta, liksom en del annat såsom parets ljudanläggning och TV.

Polis tillkallas och all den vila som Sam och Mino tankat på Seychellerna försvinner i kaoset. Det ordnas med en ny dörr och med städning, och paret inventerar vad som försvunnit och vad som finns kvar. Inalles saknas föremål till ett värde om 350 000 kronor. Fyra dagar senare har de två blivit så lugna att de åter kan tänka klart: de börjar undra varför inte det här redan är löst – skulle inte bevakningsbolaget märka sådant och skicka personal inom loppet av tio minuter?

Hur inbrottet gick till

När Mino ringer till bolaget, visar sig detta ha ganska stora bekymmer också med andra kunder. Jo, larmet fungerade. Jo, larmet kom fram den natt då inbrottet skedde. Problemet var bara att i sista sekund den dagen väldigt många anställda (18 av 40 personer) hade sjukanmält sig. Den ifrågavarande natten drabbades staden dessutom av något av en inbrottsvåg, och företagets aktiva personal räckte inte till för att ta hand om alla inkommande larm. Också polisen var överbelastad, och när bevakningsföretaget den natten försökte få kontakt med myndigheten, kom det inte fram. Företagets personal var alltså tvungen att prioritera, och tyvärr hamnade Sams och Minos lägenhet bland de larm som företaget aldrig tog om hand. Det verkar därefter ha blivit så att företaget, när det dagen därefter försökte ”städa upp” bland de larm som inte tagits om hand, förbisett larmet från parets lägenhet och inte hört av sig till vare sig paret eller polisen.

Företaget erkänner naturligtvis att det inte får vara så här. Det ursäktar sig mångordigt och erbjuder utan omsvep att det ska återbetala alla pengar som det fått för uppdraget. Mino undrar dock förstås om allt det som är borta: det ska väl företaget ersätta? Vid den punkten blir det tvärstopp: företaget menar att det inte kunde räkna med att knappt hälften av personalstyrkan

skulle sjukanmäla sig samtidigt just precis samma natt som en inbrottsvåg drabbar staden – det är inte verkligen företagets fel att de blivit så här.

Företaget har också – dock utan att på eget initiativ meddela Sam och Mino om detta, för vilket företagets företrädare ber om ursäkt – rekonstruerat inbrottet. Materialet från en övervakningskamera och samtal med grannarna visar att hela inbrottet var över på strax under tio minuter: tjuvarna var väldigt målinriktade och helt hänsynslösa (vilket förklarar den förstörda keramiken i lägenheten). Även om företagets personal skulle ha varit på plats inom rätt tid skulle stölden alltså redan ha skett när företaget reagerat. Sam och Mino borde närmast vara glada för att bevakningsföretaget inte förstörde deras semester genom att ringa och berätta om att lägenheten hemma dessvärre var plundrad. Mino och hans samtalsmotpart hos företaget bråkar om detta en stund, men till slut ger Mino upp och meddelar att paret kommer att höra av sig med rättsliga krav så snart det går: det här accepterar Mino inte!

Vatten överallt

När lägenheten väl är städad, ekar den tom. Här behövs ny inredning – vilken tur i oturen att paret just fått så många fina bröllopspresenter med vilka tomheten kan fyllas! När Sam åker för att
hämta ut några av grejorna från det hyrda förrådet för att titta närmare på dem och bestämma var i lägenheten de ska stå eller hänga, möter hon dock ganska omfattande förödelse. Själva den fastighet i vilken parets lagringsutrymme finns har nämligen, utan att någon kunde märka det tidigare, drabbats av fukt. Felsökning pågår ännu, och företaget har därför ännu ingen förklaring till hur skadan uppstått, men fukten har oturligt nog trängt igenom väggen just i den sektion där Sams bolags lager finns. Dessutom har det regnat kraftigt medan paret var borta. Kartongerna med vin som paret fått i bröllopsgåva är så fuktskadade att några flaskor ramlat ut och ned på golvet – vin värt 15 000 kronor ligger i en pöl på betonggolvet –, textiler är genomvåta (men går nog att rädda) och – värst av allt – den tavla som Sams följare målat för paret är numera en enda sörja av olja, duk och vatten.

Sam – som vid det här laget börjar bli nära gråten – säger till lagringsföretagets anställda att detta kommer att ha konsekvenser: de här skadorna ska ersättas! Den anställda svarar att företaget redan har en policy, nämligen att minska hyran för utrymmet med 50 procent. Minska hyran, undrar Sam? Det här är väl ingen hyra, utan det är lagring, och företaget har förvarat parets grejor på ett fullkomligt oansvarigt sätt! Här ska inget betalas och dessutom skadestånd utges för allt som nu är förstört! Nej, säger den anställda: paret har hyrt ett utrymme i fastigheten och utrymmet var dåligt på grund av skador i fastigheten som företaget inte kunde veta om. Det är endast detta som paret kan få ersättning för.

Notan

Det är ett ganska olyckligt nygift par som under de efterföljande dagarna försöker åter få fotfäste i sitt liv. Det känns om att hela världen är emot Sam och Mino, att inget någonsin lyckas för dem, att bröllopet var en katastrof och att de egentligen helst skulle vilja flytta. Mitt i allt elände gör Sam tre dagar efter besöket i lagringslokalen en inventering av hur mycket pengar som de två egentligen gemensamt kan få ihop innan något skadestånd utbetalats av de olika företagen. När hon surfar in på parets internetbank för att räkna ihop konton och aktieinnehav, lägger hon märke till att ett belopp på 50 000 kronor bokats från parets löpande konto: Mino glömde att vid avbeställningen av motorbåten återkalla betalningsuppdraget! Det brister för Sam och paret har nu sitt första äktenskapsgräl. På en punkt är de dock eniga: det är förbaskat konstigt att de ska behöva betala en avgift för en båt de aldrig använt. De undrar om det egentligen kan vara rätt?

När båda lugnat sig igen och med ett stärkande glas vin sitter vid köksbordet, börjar de fundera på vad de nu egentligen kan kräva.

Din uppgift

Ta på ett motiverat sätt ställning till hur de konflikter och anspråk som du identifierar kan eller ska lösas juridiskt. Inta därvid en objektiv rådgivande roll: du ska inte föras parets talan, utan på ett så objektivt sätt som du förmår upplysa det om vilka lösningar som juridiken här tillhandahåller. Du ska utgå ifrån att alla konflikter löses enligt svensk rätt.

Frågor om försäkring faller utanför kursens område och ska inte tas upp i ditt lösningsförslag.

Ju mer av berättelsen du kan hantera, desto bättre. Men kom ihåg att du samlar underlag till din portfölj och att det därför kan vara motiverat att prioritera olika delar av berättelsen.