CAN AI MAKE A CASE?
AI VS. LAWYER IN THE DUTCH LEGAL CONTEXT

Lena Wrzesniowska[*]

Abstract: The integration of AI, specifically GPT-4, into the legal field is a subject of both potential promise and intricate challenges. This thesis delves into the transformational possibilities of AI within the Dutch legal context, examining not only the quality and persuasiveness of AI-generated legal argumentation but also its competence in information retrieval, as measured by models' ability to spot relevant legal issues. An experiment was conducted with 25 legal professionals, using a real-world Dutch case with the purpose to assess GPT-4's capabilities against that of a human lawyer. To enable GPT-4 to handle case documents, the author first performed so-called co-reference resolution to remove ambiguities. Given the token limitations of GPT-4, a so-called prompt reducer technique was used to compress the text, retaining essential information. The above methods produced a coherent and full case summary within GPT-4's token constraints. This case summary, together with original lawyer's letter (denoted as Text A) was fed to ChatGPT-4 to obtain its AI-written alternative (Text B). The study subjects were presented with a case summary as well as both texts and asked for their preferences. The outcome of the experiment is as follow: 80% of participants chose the AI's composed legal document, demonstrating a strong preference for both its linguistic competencies as well its ability to spot relevant legal issues. This preference for GPT-4 writing is very consistent among genders, age groups and professions surveyed. Contextualising these findings within the broader implications for legal practice, the thesis explores potential benefits including increased access to justice and transformation of certain legal procedures. Insights and recommendations are offered for legal professionals, considering the technological evolution and ethical considerations inherent in AI integration. Acknowledging the need for further exploration, the study recognises its own limitations and encourages replication to solidify the understanding of AI's transformative role in the legal realm.

Keywords: GPT-4; AI; AI vs Lawyer; Legal Writing; Language Technology; LLM; Dutch Legal Context

[*] University of Amsterdam, Netherlands.

Table of Contents

INTRODUCTION

A. Background and Context

The advent and increasing sophistication of emerging, AI-fuelled language technologies have ignited a transformation across various sectors, including the judicial system. These technologies have the potential to revolutionise the way legal professionals work and interact with each other and the public. However, their impact on the legal system is not well understood, particularly from different perspectives such as judges, advocates, legal start-ups, and the general public, especially the poor, who traditionally, have had limited access to legal support and the pro-bono sector.

This study will assess the potential risks and issues of utilising these technologies within the current legal framework. At a broad level, the widespread use of automated case filings can congest the legal system. On a case-by-case basis, overly lengthy, automatically generated legal arguments that extensively cite a vast number of cases can significantly prolong the time required for judges to prepare and delay sentencing. Furthermore, while these technologies can generate vast amounts of legal text, the question remains: how persuasive or qualitative is this AI-produced legal writing compared to that written by humans? The efficacy of these technologies in creating compelling legal arguments is a largely uncharted territory that this study aims to explore. Thus, this research will not only investigate the impact of these technologies on the legal system but also probe their effectiveness from a qualitative standpoint, offering a multi-perspective view within the context of the Netherlands' legal landscape.

The goal is to highlight the potential risks and benefits inherent in integrating AI into the legal system and to propose recommendations for maximising efficiency while minimising adverse effects. The findings could illuminate paths to improving access to justice and streamline legal processes while maintaining a human-centric approach to law. As such, this research holds significant potential for shaping policy decisions and technological strategies in the legal domain.

B. Research Problem and Motivation

As we stand at the precipice of the AI revolution in law, a critical question looms - how will AI reshape our understanding and practice of law? Although some hope that AI language technologies will redefine efficiency and accessibility of legal services, their impact on the quality and persuasiveness of legal argumentation is uncharted territory. The problem this research seeks to address lies in this gap: how do these emerging technologies alter the landscape of legal practice, particularly in the context of the Netherlands?

The urgency of this inquiry is underscored by recent events, as AI technologies are beginning to be employed in governmental processes. A prime example is the use of ChatGPT by MPs Hind Dekker-Abdulaziz and Paul van Meenen of D66 to formulate a parliamentary motion in February 2023 (Dekker-Abdulaziz, 2023). This event highlights that these technologies have permeated into the legislative process, potentially impacting decision-making without a comprehensive understanding of the quality or effectiveness of their output. It is for this reason, why the study concentrates on the unveiling quality in the context of persuasive ability of AI-generated legal arguments, a factor which could significantly impact how law is practised in the future.

To undertake this investigation, this study conducts an experiment designed to compare AI-generated legal arguments against those formulated by human legal practitioner. Judges and advocates have been selected as participants due to their experience and knowledge in evaluating legal arguments. They will be tasked with assessing the persuasiveness and effectiveness of both a human written and an AI-generated closing argument taken from a real case.

The overarching objective of this research is to shed light on the potential impacts of AI language technologies on the legal system. By determining the quality of AI-generated legal arguments, the study hopes to provide an informed perspective on the potential speed and manner of AI integration into the legal sphere. The results are expected to offer valuable insights not only for legal professionals and technology providers, but also for the wider public who stand to be affected by this shift in the legal landscape.

C. Research Question and Objectives

The primary research question this study seeks to answer is:

How does the quality of AI-generated legal argumentation compare with human writing, and what might the implications of these differences be on the future trajectory of the legal profession in the Netherlands?

This question is pursued with the understanding that the answer will not only serve the immediate needs of the legal field, but also help shape the future development and integration of AI technologies in the broader justice system.

Research Objectives:

- To assess the quality of AI-generated legal argumentation. This will be carried out through an experiment in which legal professionals, including judges and advocates, evaluate AI-generated vs human-written legal letter. The quality of both letters will be measured in the following 4 dimensions: overall persuasiveness, clarity and coherence, strength of key arguments and use of evidence.

- To determine potential implications for the future integration of AI language technologies in the legal system. With the data collected, the study will discuss potential trajectories of AI integration in legal practice.

- To contribute to strategy development for the integration of AI in the legal system. By providing a nuanced understanding of the quality and impact of AI-generated legal arguments, this study aims to assist in the formulation of guidelines, policies, and strategies for the adoption of such technologies.

These objectives frame the scope of the research, guiding the methodology and the interpretation of the findings. The answers obtained will have implications for legal practice, technology development and societal understanding of AI's role in law.

D. Significance and Contribution

The significance of this research is manifold. Firstly, it contributes to the budding academic and professional discourse on the integration of AI in the legal profession, a field that has historically relied heavily on human expertise in linguistics, semantics, and the nuanced interpretation of legal texts. As AI models such as GPT-4 have been developed with a language-centric approach, they inherently hold potential for applications within this text-heavy field. By comparing the quality of AI-generated legal arguments with those created by humans, this study provides empirical evidence to either support or challenge the increasing use of AI in legal processes, particularly in the Netherlands.

Furthermore, this study addresses a crucial societal issue. In the Netherlands, access to traditional legal advice can be prohibitively expensive, with average hourly rates for a Dutch lawyer working on consumer law cases at around 190 EUR excluding taxes (Dutch Law, 2023) being in sharp contrast with current minimum hourly wage for adults at 11,51 EUR (Government.nl, 2023). This discrepancy, which equates to the lawyer's fees being nearly sixteen times the minimum wage, emphasises the significant financial barriers to accessing legal services for many individuals. While there is a government-supported agency (Het Juridisch Loket) offering free legal advice to aid the economically disadvantaged, their services are often slow, of poor quality (Trustpilot, ⅖ stars, 2023) and only available in Dutch. This language barrier effectively excludes a substantial expatriate population in the Netherlands, which according to the Dutch Central Bureau for Statistics, was over 800,000 in 2021 and is continually growing (CBS, 2021).

The potential of AI to generate legal arguments and advice efficiently and in multiple languages could offer a solution to these challenges. It could increase accessibility and timeliness of legal advice, and improve equity within the legal system by providing services to those who might not otherwise have access to them.

This work is also pertinent in light of the rapid advancements in AI technology and its infiltration into various sectors, including legislative processes as seen in the Dutch parliament. By examining the quality of AI-generated legal writing, this study can contribute to policy making regarding AI use in such critical areas.

Lastly, this research has implications beyond its immediate scope of application. While the focus of this study lies in the legal field, the findings can have broader relevance in the AI technology domain. Uncovering the strengths and weaknesses of AI-generated legal argumentation can indirectly contribute to the future evolution and refinement of AI language models like GPT-4 LLM. By pinpointing the areas where AI meets or falls short of human performance, this research could provide valuable insights that help guide future advancements in AI technologies for a range of professional contexts, not limited to the legal profession.

I. LITERATURE REVIEW

A. Overview of Emerging Language Technologies in the Legal System

The recent advancement in AI has brought forth impressive developments at the intersection of Deep Learning and Natural Language Processing (NLP), the most prominent of which is the GPT-4 language model developed by OpenAI. GPT-4, short for Generative Pretrained Transformer 4, is an AI model that utilises machine learning to produce human-like text. In the words of its makers:

‘One of the main goals of developing such models is to improve their ability to understand and generate natural language text, particularly in more complex and nuanced scenarios.’ (OpenAI, 2023)

This goal has a profound relevance to the legal sphere because law is characterised by a multitude of complex nuances as well as an intricate language and precise terminology. The interpretation, summarisation and application of these elements in drafting legal documents are tasks where GPT-4 could potentially be very valuable.

Katz et al. (2023) provided a compelling study on the capabilities of GPT-4 in a legal context. In their paper, ‘GPT-4 Passes the Bar Exam’, they evaluate the performance of GPT-4 against the Uniform Bar Examination (UBE), a test that includes both multiple-choice and open-ended components. The authors found that GPT-4 significantly outperformed both human test-takers and prior AI models in multiple areas, scoring well above the passing threshold for all UBE jurisdictions. This work demonstrated the potential of such models to assist in the delivery of legal services. However, these promising results should be examined with caution.

While GPT-4 and other AI models exhibit considerable skill in some areas, their performance can still present challenges in others. For instance, Savelka et al. (2023) highlight that GPT-4, while demonstrating surface-level proficiency in generating explanations of legal terms, shows limitations in the factual accuracy of the explanations it produces. They note that GPT-4 tends to ‘hallucinate’, or invent incorrect statements, underlining the need for further refinement of such technologies for reliable application in legal tasks. While the study by Savelka et al. (2023) offers interesting insights into the limitations of GPT-4 in generating accurate explanations of legal terms, it is important to note that the authors used a default temperature setting of 0.7 (on 0 to 1 scale) and this might not have been optimal for this task. This is because temperature setting is the measure of randomness of the model’s output. It is often branded as the parameter setting for creativity, because lower temperature typically decreases diversity in the model’s response. Authors themselves acknowledge that a higher temperature setting can lead to more creative but potentially less factual outputs, often branded as ‘hallucinations’. This might have contributed to the inaccuracies observed in the model’s outputs, underscoring the need for careful adjustment of GPT-4’s parameters when used in legal contexts.

Even before the emergence of GPT-4, the idea of implementing AI in the realm of law was not novel. Xiao et al. (2021) in their research titled ‘Lawformer: A pre-trained language model for Chinese legal long documents’ showcased this application. They developed *Lawformer*, a language model tailored to navigate extensive Chinese legal documents. After training the model on a vast collection of legal texts to equip it with a robust legal knowledge base, its proficiency was evaluated. Using the *Chinese Judicial Reading Comprehension* dataset, Lawformer’s responses were compared to the dataset’s annotated answers. Its performance was quantified using the *Exact Match (EM)* and *F1* score metrics, revealing significant ability in understanding long-form documents. This example serves to highlight the flexibility and potential of AI technologies in various legal and linguistic contexts across the globe.

A more recent study from China ('Legal Syllogism Prompting: Teaching Large Language Models for Legal Judgement Prediction', 2023), building on previous work in the judgement prediction domain by Katz et al. (2017), offers further insight into how large language models (LLMs) can be tailored for legal applications. Their research focuses on using AI to predict legal judgments, a critical aspect of the legal process. The authors proposed a new approach known as '*legal syllogism prompting (LoT)*' to enhance AI's performance in this field. The methodology involves teaching LLMs, specifically GPT-3 in their study, to understand and apply the structure of legal syllogism - a common form of deductive reasoning in legal analysis involving a major premise (law), a minor premise (case facts), and a conclusion (judgement). By instilling this reasoning structure, they found that the model could generate the syllogistic reasoning process of a case and provide a judgement without needing additional learning, fine-tuning, or specific examples. Most notably, this method enhanced the AI's explainability by enabling it to provide not only the final judgement but also the legal articles and justification used to reach that judgement. This study reaffirms the potential of AI technologies like GPT models in revolutionising legal tasks by offering a unique approach to predict legal judgments more efficiently and transparently.

This study aims to contribute further to evolving body of research, with a particular focus on the quality and persuasiveness of AI-generated legal argumentation in the context of the Dutch legal system.

B. Advantages and limitations of using AI in legal services

1. Advantages

Although the last official count was performed back in 2004, we know that the volume of law in the Netherlands is persistently increasing, which further complicates legal work. According to a study by De Jong and Herweijer (2004) on the development of the number of laws and ministerial regulations in the Netherlands, there has been a consistent growth in legislation since the 1970s. This growth in legislative content not only surpasses the capacity of a single legal professional, but it also necessitates the existence of various legal specialisations, as it becomes nearly impossible for any individual to maintain comprehensive knowledge of all legal domains. The proliferation of law underpins the need for machine-assisted support. GPT-4 could potentially be leveraged to address the challenges posed by this continuous expansion of law.

Incorporating AI would result in obtaining efficiency and scale, allowing legal professionals to process vast amounts of information in reduced time, thereby speeding up case handling and decision-making. Some top law firms are already looking to hire GPT Legal Prompt Engineers to help them with the integration of these technologies in their business (Hinkley, 2023), while others are developing their own GPT-based chatbots to assist with tasks such as drafting mergers and acquisition documents (Beioley & Criddle, 2023). In addition, for contracts written in Dutch, LegalFly BV, a Belgian start-up, is preparing to release an AI assistant that will facilitate aspects of contract creation, such as drafting, legal compliance, and expert guidance. According to the LegalFly (n.d.), their users will soon be able to upload and anonymise their contracts, and then have it scanned by the AI legal assistant, which highlights potential problems and weaknesses of the drafts.

Cost-effectiveness is another significant benefit, because AI can perform numerous tasks that would otherwise require substantial human labour, resulting in reduction of overall expenses. Dean Andrew Perlman proved this point by co-writing a 14-page law review article with ChatGPT in just one hour (Greene, 2022). The technology also promotes accessibility and making legal advice and support available to everyone, including those with limited resources. Furthermore, automation enables consistent execution of repetitive tasks, minimising human error, and freeing up professionals to focus on more complex and personalised aspects of legal practice, such as stakeholder management and customer service. Together, these advantages mark a transformative era in legal services, by hopefully making justice more attainable and streamlined for all.

2. Limitations and Ethical Concerns

Previous section mentioned that AI can minimise instances of human error in legal practise. However, AI has its own limitations that must be acknowledged. Firstly, GPT-4 has the tendency to 'hallucinate' - a term used to describe the generation of output that is incorrect (Schwarcz & Choi, 2023). Such hallucinations are very problematic because they can appear 'seemingly realistic' (Alkaissi & McFarlane, 2023, p3) to general audience.

Secondly, GPT-4 and LLMs in general do not have case-specific information that often informs legal strategy taken by lawyers. As highlighted in more detail in chapter 5.1, real-world legal practice requires a nuanced understanding of the client's broader circumstances as well as their risk tolerance.

Thirdly, the utilisation of AI in practice raises complex challenges concerning legal responsibility and accountability (Nolan, 2022). Unlike human legal professionals who can be held accountable for legal malpractice, AI-assisted services operate in an area where legal frameworks are often ambiguous or non-existent. This lack of clarity poses significant risks for both practitioners and clients.

Furthermore, ethical considerations and the potential for systematically perpetuating human biases are not to be overlooked. While a single biased judge may influence a finite number of cases in their lifetime, an AI model with embedded biases could affect an exponentially larger number of decisions within a short period. The implications of this increased impact are profound. A biased AI model could exacerbate existing inequalities and injustices in the legal system, affecting many more people in a day than a biased judge would in their entire career.

Finally, potential job displacement in the legal field due to the adoption of AI and automation technologies is a concern that is often brought up (Clifton et al., 2020; Helsten, 2019; Macey-Dare, 2023). However, the probable economic disruption, including job displacement, may not be as worrying as it seems. Throughout history, technological advancements have often led to positive changes for individuals and society, transforming the way we work rather than eliminating work entirely. The integration of AI into the legal profession could likewise open up new opportunities and roles, rather than simply replacing existing ones.

II. METHODOLOGY

A. Research Design and Approach

The experiment described in this work is designed using a real-world case from 2020 from the Netherlands. The case involves a legal dispute over an employment contract between Company X and Employee X, and because Employee X is an expat, both parties agreed that all the communication relating to this case was to be conducted in English. The case is represented by a collection of 10 legal documents containing arguments from both sides. For the purpose of this study, one of the final letters - comprising legal arguments made by Employee X's lawyer (document A, later to be referred as *Text A*) - was chosen as the focus of experimentation.

To enable GPT-4 to possess the same prior case knowledge as the lawyer had when composing the original document A, a method of prompt engineering was employed on the 9 remaining documents.

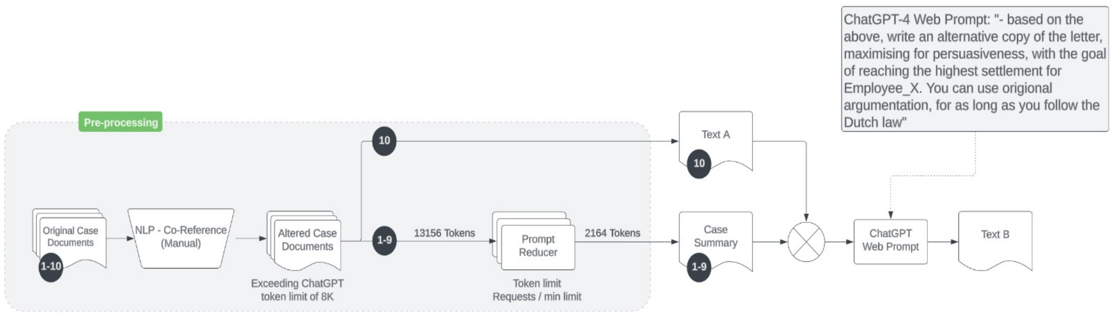


Figure 1: Text B generation - high level process flow

1. NLP - Co-Reference

Initial summarisation tests exhibited several factual inaccuracies, despite the temperature setting being set to 0. After analysis two key issues were identified: differing author perspectives across the documents and ambiguity in pronoun usage, e.g., 'my client'.

To address both issues, every document was introduced explicitly, and all ambiguous mentions were disambiguated using an NLP task known as Co-Reference Resolution[1]. This task was performed manually, by adapting the file names to include descriptive annotations to each of the 10 documents.

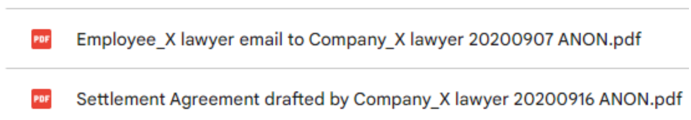


Figure 2: Manual in-file name Co-Reference Resolution examples

[1] (Ravenscroft et al., 2021)

2. GPT-4 Tokens Setup

The GPT-4 API total token limit at the time of this experiment was 8000. This number had to include the text with system instructions, GPT prompt, as well as each of the 9 documents. It was therefore decided to use the maximum of 6500 tokens towards the document, leaving 1500 tokens towards instructions.

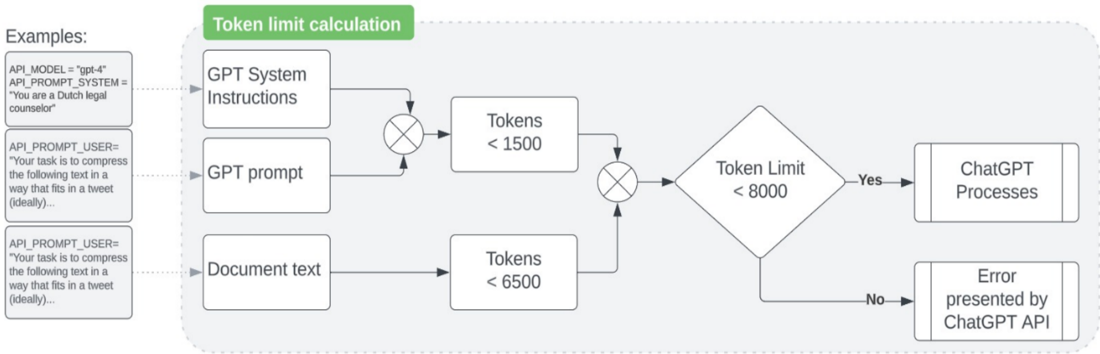


Figure 3: Token limit calculation

3. Prompt Reduction

The total length of the 9 documents needed for summarisation exceeded the token limit of GPT-4 API, hence necessitating the use of a 'Prompt Reducer' technique to compress the text.

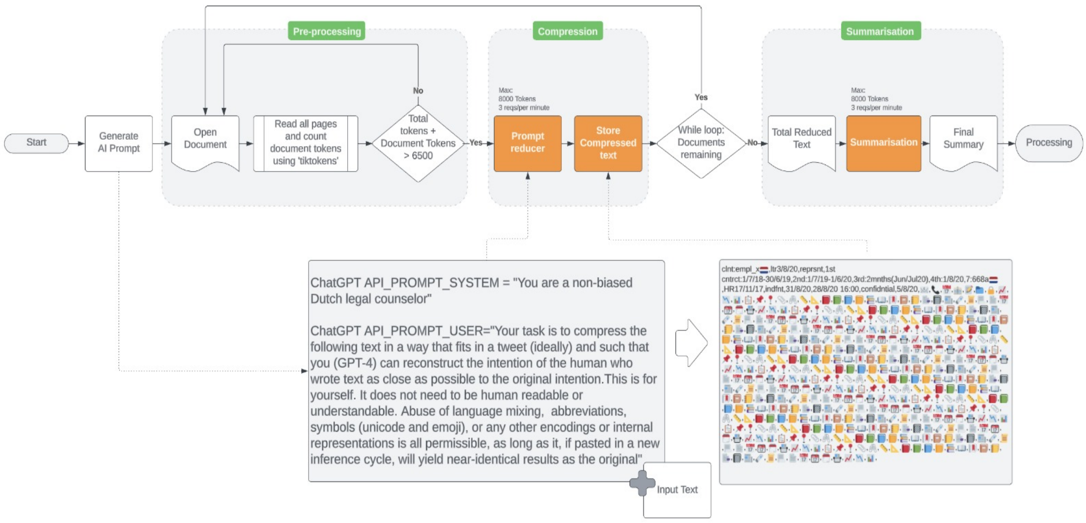


Figure 4: Pre-processing and compression process

This technique was inspired by a tweet from McKay Wrigley. The concept behind this method is to shrink the text in such a way that it is optimally concise, yet the language model can reconstruct the original intent of the text as closely as possible. This allows the language model to retain the crucial information from the text while eliminating the non-essential parts. Remarkably, the result does not need to be human-readable; it can be an amalgamation of abbreviations, symbols, unicode, and other encodings, as long as it helps the model yield nearly identical results as the original text when used in an inference cycle. Implementing this process involved a Python script

(Appendix 2) that used the OpenAI API to instruct the model to condense the document summaries. The script operated iteratively on each document, checking the total number of tokens, and if adding another document would surpass the model's token limit, it applied the 'Prompt Reducer' to the current batch of documents. This process continued until all documents were incorporated.

This approach, though novel and not yet documented in any formal literature at the time of this research, proved essential for presenting the model with a broad yet comprehensive overview of the case, enabling AI to generate effective legal document.

As a final step, an additional line (below in red) was manually added about the employee's consistent performance and bonus history, which had been mentioned in one of the case documents but not in the lawyer's original letter. This was done to test if GPT-4 could incorporate this point into its own argumentation. Final, 9 document summary used in the experiment reads as follows:

The client (Employee_X) is represented in a dispute regarding their employment contracts. They had two fixed-term contracts, followed by a third contract with a two-month extension. The client had their yearly bonuses always paid out and no registered performance issues until the contract conflict arose. The client argues that they should have an indefinite contract based on Dutch law, because they continued to work past that 2 month extension. The client is also involved in a dispute regarding the protection of confidential information. They have forwarded emails to their lawyer, but no sensitive data was shared.

Figure 5: Final case summary

The revised and compressed summary was then used as the input for ChatGPT-4 followed by the document A (in the below prompt referenced as *letter*). The prompt provided to the AI model was:

'Based on the above, write an alternative copy of the letter, maximising persuasiveness, with the goal of reaching the highest settlement for Employee_X. Use original argumentation where applicable, for as long as you follow Dutch law.'

Figure 6: Prompt used in ChatGPT-4 to obtain text B

This prompt provided to GPT-4 was specifically designed to maintain a neutral stance. The directive, *'Use original argumentation where applicable'* was deliberately devoid of explicit commands that could direct the model towards identifying and integrating arguments that were overlooked by the human lawyer, such as the aforementioned fact that the Company X had consistently paid bonuses and had no registered performance issues related to Employee X until the contract dispute arose. This strategic ambiguity was intended to evaluate whether GPT-4 would independently identify and use these untapped arguments in its response.

> *In addition to the confidentiality matter, it is important to note that Employee_X has consistently received their yearly bonuses and has had no performance issues registered until the contract conflict arose. This further supports the notion that our client has been a valuable asset to your client's company, which should be taken into consideration when assessing the settlement.*

Figure 7: GPT4 spots and includes the additional argument into text B.

As demonstrated by the outcomes - the above excerpt comes from the GPT-4 generated text B - the AI model effectively discerned and utilised these neglected points, thereby attesting to its robust information retrieval capabilities within the legal context.

B. Data Collection and Analysis Methods

The primary method of data collection employed in this study was an online survey designed on Google Forms, titled ‘Legal Writing Research.’ The survey included several components designed to provide comprehensive insights into the study’s key objective: comparing human-generated and AI-generated legal arguments.

The dimensions along which the effectiveness of legal writing was assessed are as follows:

- **Persuasiveness** - Assesses the ability of the text to convince the reader of its stance.
- **Clarity & Coherence** - Examines how well the text is structured and how easily it can be understood.
- **Strength of Key Arguments** - Evaluates the robustness and validity of the core arguments presented.
- **Use of Key Evidence -** Measures information retrieval capacity, a crucial aspect that extends beyond mere stylistic attributes such as persuasiveness, clarity, and coherence. While these linguistic elements are essential for constructing a compelling legal argument, the integration of accurate, relevant, and compelling evidence that supports the argument being made in a legal document is also fundamental. It involves not only spotting the right evidence from large bodies of text but also understanding how to utilise that evidence effectively within the argument. This process requires a deeper analytical skill that goes beyond mere linguistic ability. Figure 8. demonstrates full structure of the survey inclusive of the dimensions just discussed.

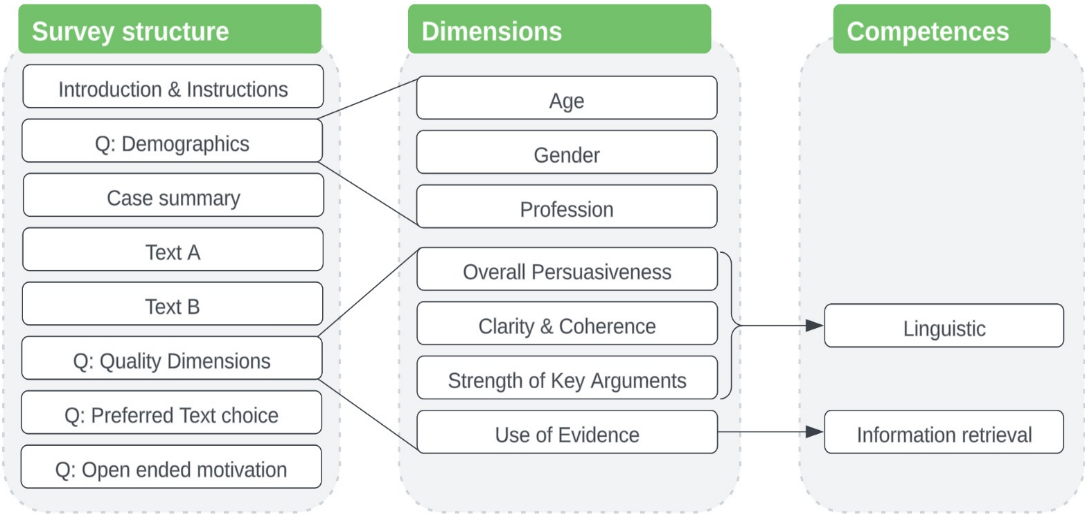


Figure 8: Survey structure

The survey (Appendix 1) began with an overview of the tasks they would undertake. It further included demographic questions, asking the participants about their age, gender, and profession. The main section of the survey comprised an earlier constructed case summary followed by two anonymised legal letters (designated as Text A and Text B), both arguing in favour of Employee X. Unbeknownst to the participants, Text A was the original legal letter penned by a human lawyer (earlier in this work referred to as document A), and Text B was an AI-generated alternative produced by GPT-4.

Participants were requested to scrutinise both Text A and Text B, after which they were asked to answer a series of questions aimed at gauging their perspectives on several facets of each letter. They rated the overall persuasiveness, clarity and coherence, and the strength of the key arguments on a scale of 1 to 10. They also evaluated the appropriateness and effectiveness of the evidence presented in the two texts. Lastly, they were asked to choose between Text A and Text B based on their overall effectiveness in presenting the case and to justify their selection.

C. Study Participants and Sampling Strategy

The distribution of the survey was conducted in a carefully phased manner. Initially, it was shared with a select group of 16 judges, whose contact information was sourced through a personal contact. The invitation email was strategically neutral, broadly mentioning language technology but deliberately avoiding any mention of GPT technology to prevent the introduction of bias. Following this, the survey was disseminated among the author’s extended professional network to reach out to lawyers and other legal professionals, including paralegals and legal assistants, to ensure a balanced dataset. Notably, while the survey was conducted in English, it was a prerequisite for all participants to be based in the Netherlands, given that the original case was based on Dutch law. The data collected from the survey was subsequently analysed to draw insightful conclusions about the comparative persuasiveness and effectiveness of human vs. AI-generated legal arguments.

The data collection phase took 8 weeks and yielded a total of 25 responses, out of which 9 were judges, 9 were lawyers, and the remaining were other legal professionals. Of note, the pool of judge participants included two members of the Supreme Court of the Netherlands (De Hoge Raad Der Nederlanden), a detail that adds a high level of expertise and authority to the responses. These identities remain anonymous in the data analysis and results reporting, maintaining the study's ethical considerations.

D. Evaluation Criteria and Scoring System

The scoring system utilised in this study was designed to quantify participants' evaluations of two legal texts - Text A and Text B. Respondents were asked to score each text on four key dimensions: overall persuasiveness, clarity and coherence, strength of key arguments, and the appropriateness and effectiveness of the evidence used. Each of these dimensions was rated on a scale from 1 to 10, with 1 being the lowest possible score (not persuasive at all/not clear at all/etc.) and 10 being the highest possible score (extremely persuasive/extremely clear/etc.). This type of Likert scale is commonly used in research to gauge respondents' attitudes or perceptions towards a particular subject (Robinson, 2014).

Responses to each of these scoring questions were then analysed both individually and in aggregate to gain insights into respondents' evaluations of the two legal texts. Average scores were computed for each text on each of the four dimensions, and these averages were compared to identify potential differences in perceived quality between the lawyer-written text and the GPT-4 written text.

In addition to the Likert scale questions, the survey included an open-ended question that asked respondents to choose which text they thought was overall more effective and to explain their choice. Responses to this question were categorised based on the chosen text (Text A or Text B), and the explanations were analysed qualitatively to identify common themes or recurring arguments.

This approach allowed to not only obtain a numerical representation of the perceived quality of the two legal texts but also to understand the reasons behind the ratings, adding depth to the analysis. By comparing these scores and the qualitative feedback, research aimed to get a comprehensive understanding of how legal professionals perceive the quality of legal writing generated by an AI like GPT-4 compared to traditional, human-generated legal writing.

E. Ethical considerations

In conducting this experiment, several ethical considerations were carefully observed:

- Informed Consent and transparency - all participants were informed, albeit in broad terms, of the purpose of the study

- Privacy and Anonymity - to respect the privacy of the participants, no personally identifiable information was collected in the Google Form except their professional role. Participants were identified solely by their profession (judge, lawyer, etc).

Email addresses used to send the Google Form were not stored or linked to the responses.

- Data Security - the data collected through Google Forms is secured by Google's privacy policies. Only the researcher has access to the responses, and the data will not be used for any purposes outside of this research.

- Bias Prevention - to mitigate bias, the details about the specific AI used in generating one of the texts were not disclosed to the participants. This was done to ensure that responses were based on the quality of the legal texts rather than preconceived notions about AI or any of the OpenAI products which were heavily discussed in mass media in the time of conducting this research.

- Post-Research Interactions - after the completion of the study, some participants reached out via email expressing their curiosity about the study's purpose and what their choices signified. While these interactions could potentially reduce the anonymity of the participants in the researcher's perspective, no further data was collected during these exchanges. Moreover, these post-study interactions did not influence the analysis and interpretation of the data already collected.

III. RESULTS AND ANALYSIS

A. Data Pre-Processing

During the analysis phase, it was noted that two of the responses were provided in Dutch, despite the questionnaire being designed fully in English. To maintain consistency in the analysis, these responses were translated into English, ensuring the sentiment and specific terminology were accurately preserved. This allowed for a comprehensive and uniform evaluation of all collected data.

Furthermore, the open-ended question in the questionnaire had a minimum character requirement set at 200. To comply with this, some respondents extended their shorter answers with filler characters (e.g., 'xxxxxxxxxxxxxxxxxxxxxxx') to meet the requirement required to submit their response. These extraneous characters were removed during the pre-processing stage to ensure a clean and accurate analysis of the responses.

B. Descriptive Statistics

The questionnaire for this experiment was designed to evaluate and compare the quality of legal writing produced by a trained lawyer and by the GPT-4 language model. It comprised two primary components: demographic questions and legal text evaluation questions.

The demographic questions were designed to gain a snapshot of the respondents' background. These questions gathered information about the respondent's age, gender, and professional background. Understanding these factors helped contextualise the feedback and control for potential biases or perspectives unique to certain demographic or professional groups.

1. Age

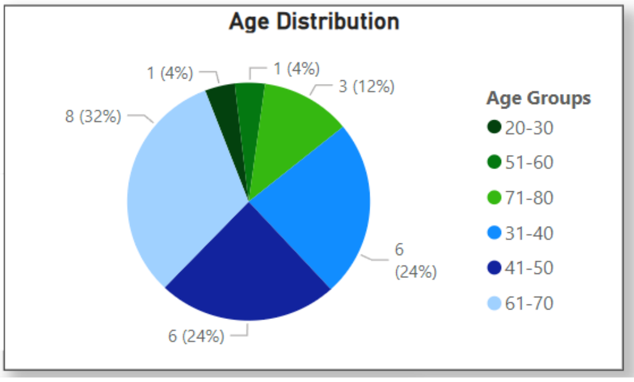


Figure 9: Age distribution

The average (mean) midpoint age is 53 and it indicates that the centre of the distribution of ages in the sample is approximately 53 years[2]. The standard deviation of 15 shows a relatively high degree of dispersion around the mean age. This indicates that there is a significant spread in the ages of the participants in the study. This diversity in age suggests that the findings may be more generalisable across different age groups.

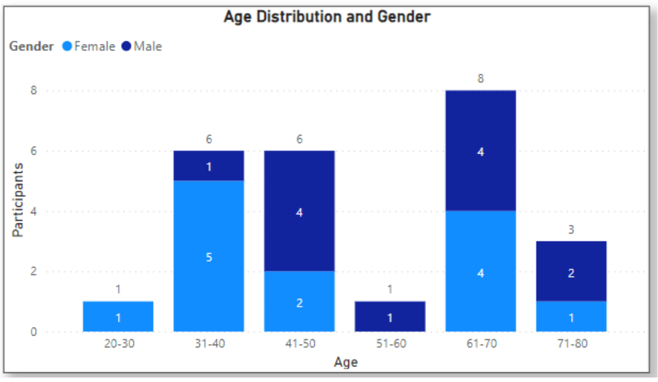


Figure 10: Age distribution correlated with Gender

The analysis reveals a balanced gender representation across age groups (Figure 10). Further examination also indicates a pronounced presence of individuals in the 'judge' profession within older age brackets (61-70 and 71-80). Conversely, the younger age groups demonstrate a stronger representation of 'lawyers' and other legal professionals.

[2] This is the average age of the participants considering the midpoint of each age range, which is an approximation of the average age of the participants.

2. Professional Background

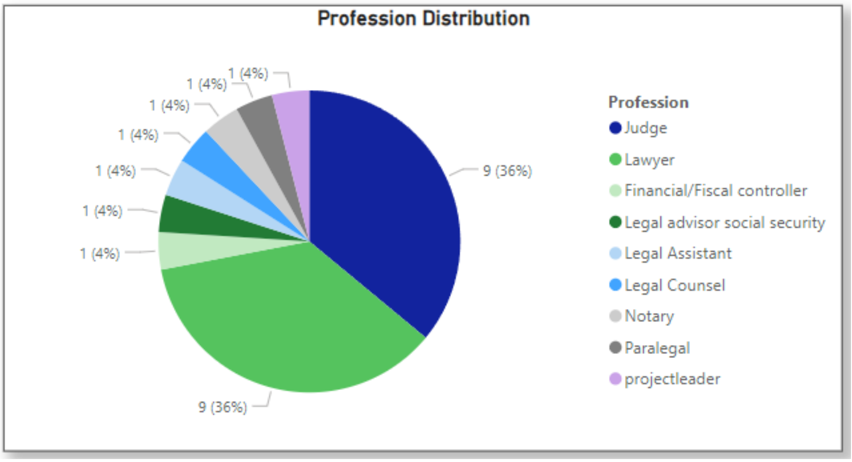


Figure 11: Professional distribution

All participants in the study have received legal education and/or training and are based in the Netherlands. However, not every participant is currently engaged in the direct practice of law. Diversely, one participant is employed as a financial controller, and another serves as a project leader within a legal department.

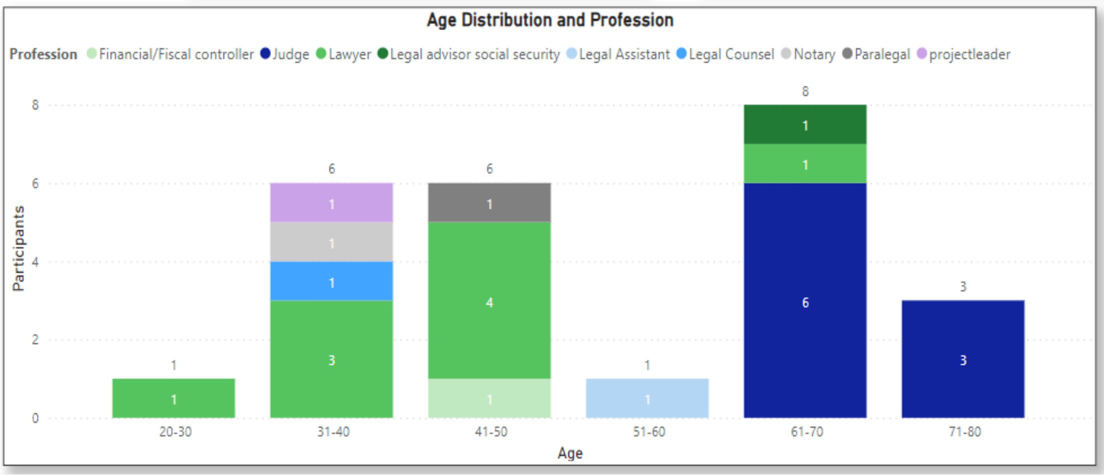


Figure 12: Age Distribution correlated with Profession

The participant pool exhibits a diverse range of expertise levels, including high-profile, senior Dutch judges. This prominent level of expertise is counterbalanced by the inclusion of participants from a variety of legal roles, such as legal assistants, a young legal counsel, and a local notary. This diversity in the professional background is intentional and provides a comprehensive perspective on the research question. A cross-analysis of the data reveals distinct age trends within professional groups. The judges participating in the survey fall into more senior age groups while the lawyers demonstrate a strong representation in the younger age brackets. This discrepancy in age distribution between judges and lawyers, while evident, was an inherent factor within the available participant pool and was beyond the scope of adjustment for this research. Recognising this disparity is important as it may introduce potential biases or unique perspectives associated with different professional experiences and generational viewpoints. Despite this imbalance, the diverse age and professional range enrich the study by providing a wide spectrum of legal expertise and perspectives.

3. Gender

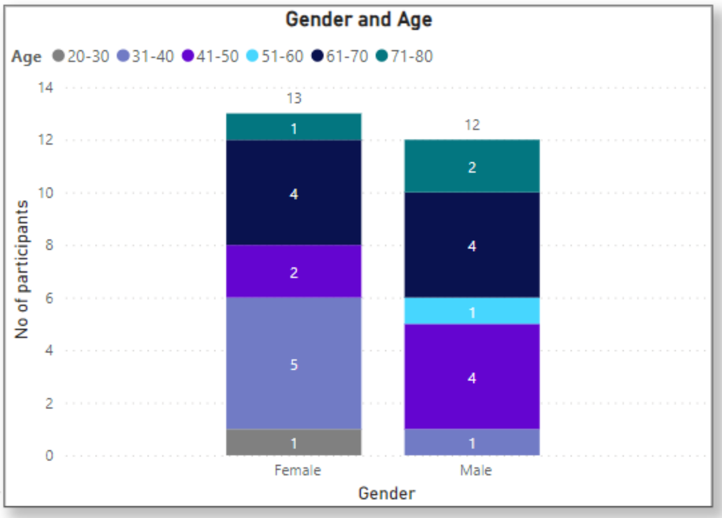


Figure 13: Gender Distribution correlated with Age

The study attained a balanced gender distribution with a near equal count of 12 male and 13 female participants. In examining the correlation of gender with age groups, a similar uniform distribution was observed.

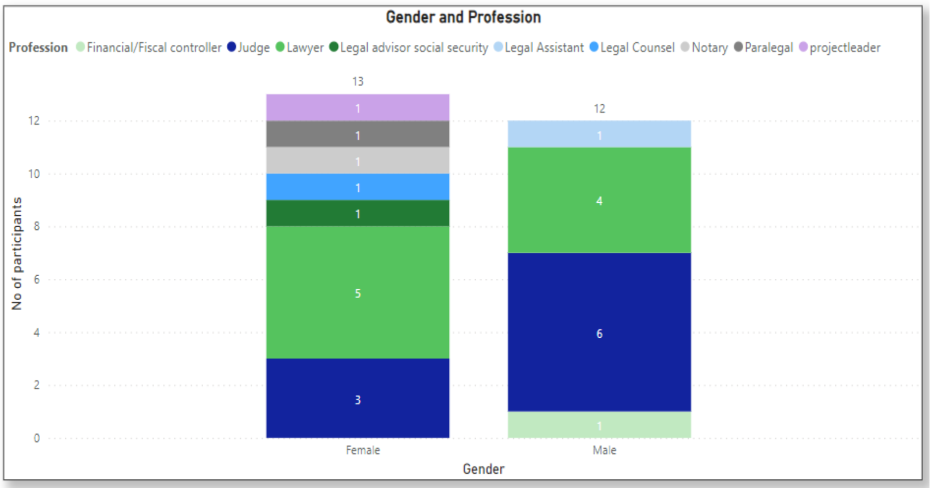


Figure 14: Gender correlated with Profession

However, an intriguing variance was identified when looking at the professions, specifically within the subset of judges: the study included three female judges compared to six male judges, providing a point of diversity within the otherwise evenly distributed dataset[3].

C. Response Analysis

1. Quality Dimensions Questions - Results

The survey (Appendix 1) asked the participants to answer 4 questions on 1-10 scale, 3 of which were designed to assess linguistic capacities (persuasiveness, clarity & coherence, strength of key arguments), while the 4th one was measuring information retrieval capacity (use of key evidence).

The below data is organised into 3 tables, each focusing on different demographic characteristics of the participants: age, profession, gender. Within each table, individual average scores are provided for Text A (human lawyer) and Text B (GPT-4), for each specific demographic group and quality dimension. For instance, participants in the Age group 20-30 have scored Text A's persuasiveness, on average, a 5 while Text B received an 8. The totals at the bottom of each table represent weighed averages. These are calculated by taking the average score for each demographic sub-group, then multiplying it by the percentage of total responses that sub-group represents. The sum of these adjusted averages across all categories forms the overall weighted average (total). This method ensures that each sub-group's contribution to the overall weighted average reflects its proportion of the total responses, thereby creating a balanced representation of the entire participant group's evaluations. Since the totals are based on the same overall set of responses, they remain consistent across each table – e.g., the total weighted average for persuasiveness dimension for Text A is 6.28 across age, gender and profession.

[3] According to the Central Bureau Statistics, 63% of judges in the Netherlands are women (CBS, 2019). The gender trend observed in this study is thus not representative of the overall gender distribution within the judge profession in the Netherlands.

Age	Avg Persuasiveness A	Avg Persuasiveness B	Avg Clarity & Coherence A	Avg Clarity & Coherence B	Avg Arguments Strength A	Avg Arguments Strength B	Avg Use of Evidence A	Avg Use of Evidence B
20-30	5.00	8.00	5.00	9.00	6.00	8.00	5.00	8.00
31-40	6.67	8.50	7.00	8.17	6.33	8.50	6.50	8.00
41-50	5.83	7.67	5.67	8.33	6.00	7.67	5.67	6.83
51-60	7.00	9.00	5.00	9.00	7.00	8.00	5.00	6.00
61-70	6.25	6.88	6.63	7.25	6.13	7.13	6.25	6.75
71-80	6.67	7.33	6.67	8.00	6.67	7.00	7.00	6.67
Total	**6.28**	**7.64**	**6.36**	**7.96**	**6.24**	**7.64**	**6.16**	**7.08**

Gender	Avg Persuasiveness A	Avg Persuasiveness B	Avg Clarity & Coherence A	Avg Clarity & Coherence B	Avg Arguments Strength A	Avg Arguments Strength B	Avg Use of Evidence A	Avg Use of Evidence B
Female	6.38	7.92	6.77	8.00	6.31	8.08	6.38	7.46
Male	6.17	7.33	5.92	7.92	6.17	7.17	5.92	6.67
Total	**6.28**	**7.64**	**6.36**	**7.96**	**6.24**	**7.64**	**6.16**	**7.08**

Profession	Avg Persuasiveness A	Avg Persuasiveness B	Avg Clarity & Coherence A	Avg Clarity & Coherence B	Avg Arguments Strength A	Avg Arguments Strength B	Avg Use of Evidence A	Avg Use of Evidence B
Financial/Fiscal controller	7.00	8.00	7.00	8.00	6.00	7.00	6.00	7.00
Judge	6.33	7.22	6.56	7.56	6.33	7.11	6.44	6.78
Lawyer	5.78	7.67	6.00	8.44	6.11	7.89	5.78	7.33
Legal advisor social security	8.00	6.00	8.00	7.00	7.00	8.00	7.00	6.00
Legal Assistant	7.00	9.00	5.00	9.00	7.00	8.00	5.00	6.00
Legal Counsel	7.00	9.00	7.00	8.00	7.00	8.00	7.00	8.00
Notary	5.00	8.00	6.00	7.00	3.00	8.00	5.00	6.00
Paralegal	6.00	8.00	6.00	8.00	6.00	8.00	6.00	8.00
projectleader	8.00	9.00	7.00	8.00	8.00	9.00	8.00	9.00
Total	**6.28**	**7.64**	**6.36**	**7.96**	**6.24**	**7.64**	**6.16**	**7.08**

Figure 15: Quality dimension averages for Text A (human lawyer) vs Text B (GPT-4)

A striking consistency emerged from the data: Text B's weighted average total scores are higher across all professions, genders and ages. Additionally, GPT-4 outperformed Text A, written by a trained lawyer, in effectively all 17 demographic sub–groups[4] and across all 4 evaluation dimensions - persuasiveness, clarity & coherence, strength of arguments and aptness of evidence use. Participants viewed the AI-generated content as superior writing, a unanimous inclination that not only strengthens the validity of the results but also emphasises the impressive capacity of AI in producing compelling legal text.

Having said that, the gaps in scores between Text A and Text B were not uniform across all dimensions. The largest gap was observed in clarity & coherence, where Text B showed a substantial lead over Text A, with a gap of 1.6 points (7.96 vs. 6.36). The smallest difference was in the use of evidence, where Text B led by only 0.92 points (7.08 vs. 6.16). That is interesting considering that Text B included evidence that Text A's lawyer failed to spot.

When examining the same data through the lens of line trends, additional key observations emerge (see Figure 16):

- Younger participants perceived a more significant difference in persuasiveness between the two texts compared to their older counterparts.
- Lawyers noted a larger gap in persuasiveness between the two texts, strongly favouring GPT-4. Judges also favoured GPT-4, but their gap in persuasiveness scores between the two texts was smaller in comparison to the lawyers' assessment.

[4] Two minor exceptions apply: the legal advisor's professional sub-group gave higher average scores to Text A in 3 out of 4 dimensions, but there was only one respondent in that sub-group. Additionally, the age sub-group 71-80 preferred Text A over Text B in the Use of Evidence category.

- In terms of clarity and coherence, lawyers represented the biggest gap in scores.[5]

- With regard to both information retrieval (Evidence Use) and Strength of Arguments, we observe that the gap of scores between the texts narrows with seniority (Age) of participants.

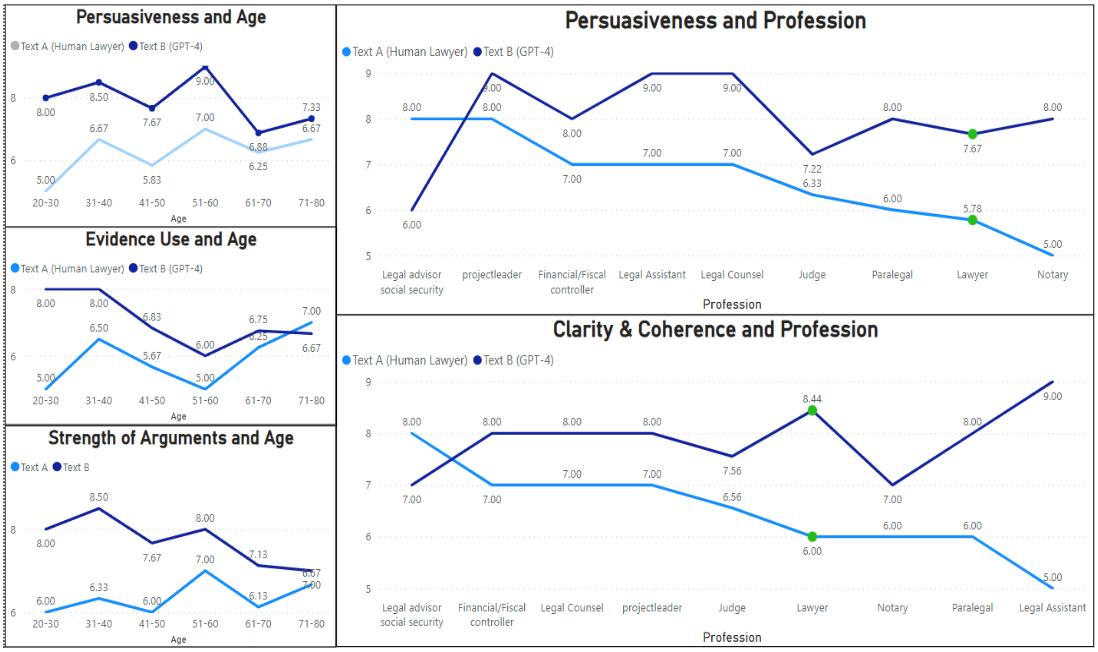


Figure 16: Line trends on selected demographic-quality dimension pairs

- Finally, when analysing Gender data, it was established that men have scored both texts lower than women across all 4 quality dimensions.

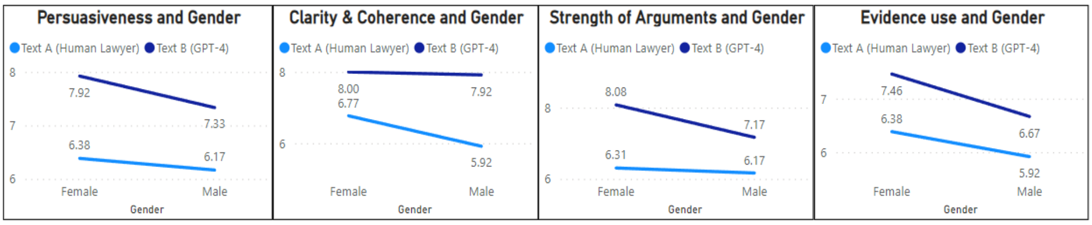
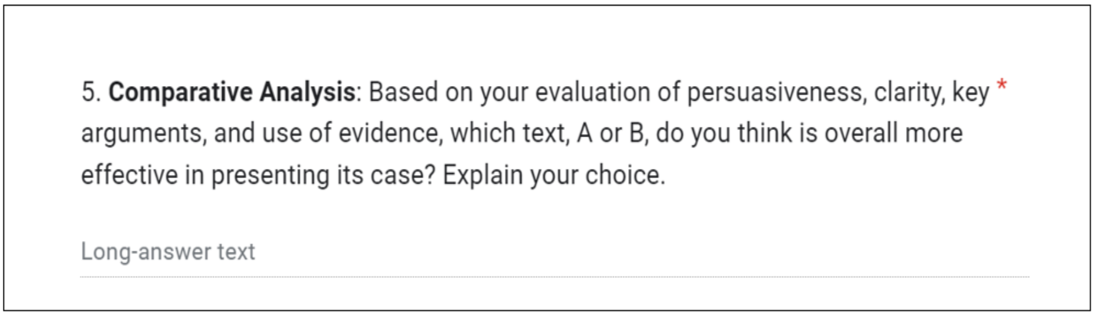


Figure 17: Gender vs quality dimensions

2. Text Choice – Results

After completing the questions related to various dimensions, the participants were required to make their final text choice by answering the following question (see Figure 18).

[5] Legal Assistant score gap is bigger, but it comprises of one respondent and therefore deemed not representative as a profession group overall.


5. **Comparative Analysis**: Based on your evaluation of persuasiveness, clarity, key arguments, and use of evidence, which text, A or B, do you think is overall more effective in presenting its case? Explain your choice. *

Long-answer text

Figure 18: Questionnaire's last question- which text is more effective?

Given the results observed in individual quality dimensions questions, it is not surprising that final text choice was overwhelmingly Text B. A substantial majority ofparticipants, a full 80%, expressed a preference for the text generated by the GPT-4 model over the document crafted by a trained lawyer.

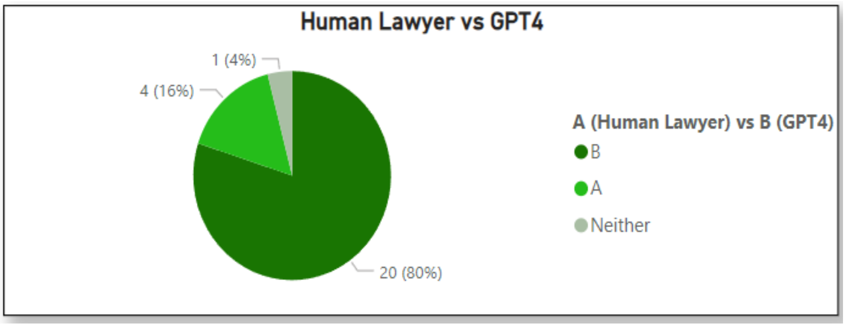


Figure 19: Final text choice responses: Text A vs Text B

The analysis further suggests a noteworthy correlation between age distribution and the final text choice. Specifically, all participants favouring the human writer fell within the 61-70 and 71-80 age groups, hinting at a potential age-related preference for human-crafted legal texts over those generated by AI. Notably, gender balance was maintained across both text preferences - of the four favouring the human lawyer's text, both genders were equally represented. Similarly, in the larger group preferring the GPT-4 text, there was an even split with ten men and ten women.

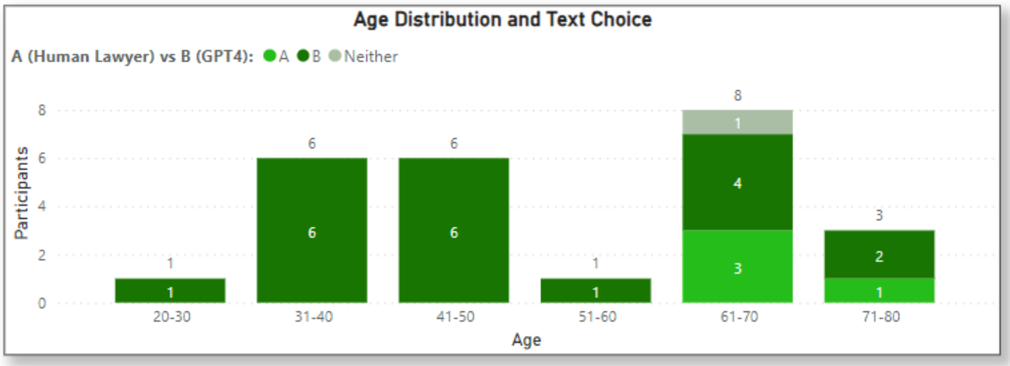


Figure 20: Age correlated with text choice

Upon reviewing the preferences for the human-written Text A, it emerges that three out of the four respondents opting for this text are judges, while one is a senior legal advisor. It is very remarkable that, despite having nine lawyers participating in the study, none expressed a preference for the human lawyer's writing. This intriguing pattern suggests potential distinctions in text preferences across different legal roles.

The analysis of the open question answers is focused on understanding the motivation behind respondents' choice and has found as follows.

3. Text Choice - Motivations

Analysis of open question responses to the final choice of text results lead to the following observations:

❖ Tone and Style

Most of the respondents have commented on the tone and style of the texts. Text B, generated by GPT-4, is described as 'more friendly' (4[6]), 'kind' (1), 'considerate' (1), 'balanced' (1), 'clear' (6), 'concise' (1), 'polite' (1), 'non-invasive' (1) and easier to read (1). Participants found this style more 'persuasive' (6) and 'pleasant' (1). On the contrary, Text A, written by the human lawyer, was often seen as 'formal' (2), 'aggressive' (1), or 'defensive' (1), and some respondents mentioned that this tone seemed to mask weak arguments or made the text harder to understand (2).

❖ Clarity and Structure

Participants appreciated the clear structure (3) and to-the-point (6) nature of Text B. It is less cluttered (2) and doesn't wander into 'unnecessary details' (1). Text B was also appreciated for better 'choice of sentences' (1) and better at 'presenting' the case (2). Text B is also said to have a 'better layout' (2). Figure 20. demonstrates a sample response which references both tone and structure of Text B as related to respondent's perceived persuasiveness of the text.

[6] Number of respondents who used each of the phases

'Text B. It's the tone of voice for me. Text A focusses on responding on the statements of the other party while Text B focusses on 'be reasonable' and tell your client to follow. As a lawyer I do wonder whether 8 months would be reasonable, but because the way they present it, it almost seems like it is.
Text B is more structured than Text A in my opinion. Because they focus on getting their point across instead of responding to the arguments of the other party it seems to have more structure and thus more persuasive.'

Figure 21: Illustrative response referring to the tone and structure of Text B and its contribution to perceived persuasiveness.[7]

❖ Argumentation and Information retrieval (Use of Evidence)

Some respondents appreciated the quality of argumentation in Text B (4), especially the clarity and persuasiveness. 1 participant considered Text A's arguments more complete, but interestingly, despite this, they still chose Text B as their preferred one. Most respondents (3) said the evidence was better in Text B, and 1 said Text A presented stronger evidence.

Importantly, an objective assessment reveals the additional argument present in Text B, not included in Text A (this argument was also mentioned in case summary provided for respondents):

'The client had their yearly bonuses always paid out and no registered performance issues until the contract conflict arose.'

Figure 22: Excerpt from Case Summary

This argument was overlooked by the original lawyer responsible for Text A but was identified and included by GPT-4 in Text B proving its ability to act as effective information retrieval tool. One respondent specifically recognised this differential evidence factor as one of the reasons for their preference of Text B:

'Text B because 1) it has more evidence (e.g. that the Employee has always performed well and received the bonuses) compared to text A. 2) text A includes unnecessary details. 3) Furthermore, text B persists on the initial amount and does not reduce further the amount of the amicable settlement unlike text A.'

Figure 23: Quote from one of the respondents – motivation for final Text Choice

---

[7] Author's Note: In this context, '8 months' refers to the length of the settlement offer proposed in Text B, representing a pay-out equivalent to 8 months' salary for Employee X.

This commentary suggests that additional or more comprehensive evidence contributes significantly to participants' perceptions of effectiveness of the text.

- Professionalism: There are mixed views on which text appears more professional. Some respondents found Text B more professional due to its tone, clarity, and structure. Others felt that the professional nature of Text A, despite its complexity and aggressiveness, made it more suitable for a legal context.

The above themes provide insight into the preferences and motivations of the study participants. They value clarity, a friendly tone, directness, resolution-seeking, and a structure that aids understanding, all of which were attributes they found more prominently in Text B. The original legal text (Text A), despite its thoroughness and professional tone, was seen as aggressive, complex, antagonistic and harder to understand, which detracted from its effectiveness in the eyes of the respondents.

What stands out is that for those favouring Text A, the preference was often described in modest terms. The differences were articulated as 'I prefer lightly Text A,' 'I find Text A more effective but only just', or that Text A 'feels better'. These statements suggest that even those 16% of participants who favoured the original lawyer's work did not have a strong or emphatic preference towards it. The subtlety in their choice indicates a level of ambivalence and highlights the closely matched quality of the two texts. Moreover, the statement that Text A 'feels better' may also reflect a tendency towards familiarity. The participant's preference might be influenced by their training and prolonged exposure to this specific form of legal writing, causing them to gravitate towards what feels more traditional or what they are more accustomed to. To fully understand this aspect, a more comprehensive investigation into how legal writing is traditionally taught in the Netherlands would be necessary. This would entail examining the writing conventions, and stylistic norms that shape the legal writing practices within the country. Such an inquiry could provide insights into why some respondents may have gravitated toward the familiar structure and tone of Text A. However, this investigation falls outside the scope of this paper.

4. Demographic Variances - Motivations

After examining the arguments made by all respondents, the study looked at how these responses differed between the main respondent groups (judges vs lawyers). Below table present these findings:

Aspect	Judges	Lawyers
Formality and language style	Prefer the formality and traditional legal style of Text A.	Appreciate the clear and concise language of Text B, finding it more understandable and potentially persuasive.
Content and persuasiveness	Acknowledge the persuasiveness of Text B due to completeness and strength of arguments.	Emphasise the structure, clarity, and non-antagonistic tone of Text B as key to persuasiveness, especially for non-lawyers.
Case specifics and evidence	Prefer the case-specific approach in Text B.	Find the additional evidence provided in Text B (consistent performance and bonuses) more compelling and persuasive.
Consistency and clarity of proposal	Point out inconsistencies in Text A's proposal.	Appreciate the clarity and persistence of Text B's proposal, particularly its refusal to further reduce the proposed settlement.
Relevance and extraneous details	No specific pattern noted.	Appreciate that Text B does not discuss the relevance of other cases and avoids unnecessary details, finding these attributes beneficial to the persuasiveness and effectiveness of the text.

Figure 24: Text Choice motivation: Judges vs Lawyers

Additionally, when analysing the female vs male respondents, the following 2 main differences were noted:

- ❖ Detail-oriented vs big picture: female respondents tended to provide more detailed feedback on specific aspects of the texts (such as formality, structure, and evidence presented), while male respondents appeared to focus more on the overall effectiveness and clarity of the communication.

- ❖ Emphasis on tone and style: female respondents seemed to give more consideration to the tone and style of the text, such as its friendliness, politeness, or aggression. In contrast, male respondents showed more interest in the text's directness and clarity.

Finally, findings also suggest that the younger age groups (20-40) favour clear, professional communication with formal language and strong structure. On the other hand, the older groups (61-80) showed a stronger inclination towards consistency, brevity, and the inclusion of more legal references. The mid-range group (41-50) exhibited a preference for a more resolution-seeking tone.

However, when looking at responses and age groups it should be noted that previously compiled age-profession analysis shows that all the respondents in the age group 71-80 are judges. Hence, this age group's preference for more legal references and their focus on specific weaknesses in argumentation might not be representative of this age group as a whole, but rather indicative of their professional background. Judges often draw on prior legal opinions and case law in formulating their own judgments, which might explain this specific preference. Therefore, the findings regarding this age group should be interpreted with this context in mind. And perhaps, the decades of experience possessed by these judges allow them to perceive nuances and complexities that may elude junior lawyers. The collective knowledge represented by LLM models such as GPT-4 may still differ from the wisdom and insight that comes with individual expertise and decades of hands-on experience. This understanding of law, sharpened over time, allows senior practitioners to see beyond mere text and delve into the underpinning legal principles and precedence. Therefore, the findings regarding this age group should be interpreted with this rich context in mind, recognising the valuable contribution of experiential learning and the profound depth of understanding that comes with years of practice in the legal field.

This distinction, be it referred to as knowledge vs wisdom, may also relate to understanding and empathising with client's special circumstances – subject to be covered in the next chapter.

IV. DISCUSSION

A. Human Factor

While the results achieved by text B are clearly impressive it is crucial to underscore one key area where human expertise continues to demonstrate its unique value: contextual understanding. The human lawyer who wrote Text A had access to information that was not explicitly stated in the legal documents. They knew, for instance, that Employee X was dealing with uncertainty caused by the global Covid-19 pandemic and its accompanying widespread job loss of 2020. The client sought to maximise their pay-out and steer clear of costly court proceedings or even long negotiations because they did not have legal insurance covering lawyer's costs.

This critical context informed the lawyer's strategic decision to lower the final settlement offer to €25K (as compared with the €38K amount suggested in Text B by GPT-4). This tactic was aimed at encouraging a swift settlement and minimising legal expenses for the client. Such a comprehensive understanding of the client's broader circumstances and risk tolerance was beyond the AI's reach, as it was not explicitly stated in the legal documents GPT-4 was trained on. Even though the participants in this study found the AI's output more persuasive overall, it is important to acknowledge that real-world legal practice often requires a nuanced understanding of their client situation that extends beyond the confines of legal documents. This discovery showcases the invaluable role that human legal professionals continue to play, even amid the rapid advancements of AI technology.

Nevertheless, it is worth mentioning that it is already possible to incorporate detailed 'special circumstances' annotations or supplementary contextual inputs. Paralegals could systematically craft these documents, ensuring their inclusion during the model's input preparation phase. By embedding such nuanced details, the GPT stands poised to produce results that resonate more deeply with the distinct intricacies of each client's situation. The effectiveness of this integration, however, remains an intriguing area for further research.

B. Contextualising the Findings

The outcome from this research project carries profound implications not just for the Dutch legal system, but also potentially for the future trajectory of legal practice globally.

GPT-4 could aid faster and more cost-efficient case preparations, especially in the preliminary stages where vast amounts of information need to be sifted through and summarised. There is also hope for the most economically disadvantaged and younger populations. The ability of AI to generate high quality legal content can potentially democratise access to good legal representation for all. This is especially important in countries such as the Netherlands, where legal expenses can be prohibitive.

Recognising Netherlands' history of swift technological adoption, it is plausible to anticipate that many upcoming legal start-ups will utilise AI-powered tools. Such platforms might employ user-friendly, question-guided interfaces, with GPT models formulating the legal advice responses in the background.

Historical patterns also suggest that these start-ups may emerge before comprehensive regulations are in place. Such a premature emergence poses potential challenges. For instance, if in a situation of legal conflict one party leverages AI and thereby gains a significant advantage in cost and speed, it could overwhelm its opponent with prohibitive legal expenses. And if the private market adopts this AI-aided methods of conflict resolution quicker than the government, and both parties resort to AI-assisted methods before the public sector can adapt, the judicial system might be overwhelmed with a plethora of motions and requests, potentially leading to administrative gridlocks resulting in delays in resolutions.

One viable solution could be for parties to stipulate, at the onset of a contract, a commitment to engage with AI-driven mediation. This would involve a neutral, third-party AI model to which each side presents their arguments—also crafted with the assistance of AI. Such an AI-enhanced mediation process would offer cost and speed benefits not only for the involved parties but also for taxpayers.

If that happens, the role of lawyers and legal professionals will inevitably evolve. Traditional tasks and responsibilities, particularly those that are routine or data-intensive, may become automated, rendering some current legal roles obsolete. However, this does not spell the end for the legal profession; instead, it underscores a need for evolution.

CONCLUSION

A. Summary of the Research Question, Findings and Limitations

The primary question of this thesis was: what is the transformational potential of AI, specifically GPT-4, in the Dutch legal context, when pitted against a human lawyer in terms of constructing effective legal arguments? The outcome of the experiment was unequivocal. Of 25 participating legal professionals, 80% exhibited a preference for the GPT-4 composed legal document over its human-authored version. This significant tilt towards the AI-generated content remained consistent across diverse demographic groups and was observed in all quality metrics related to both linguistic competences as well as information retrieval abilities.

While these findings are compelling, it is essential to acknowledge the inherent limitations of any single study. To fully assess the robustness and generalisability of these results, especially in light of the ever-evolving nature of GPT models, replication by other independent researchers in varied settings is recommended. This would provide a broader perspective and further validation of the transformational potential of AI in the legal realm.

B. Recommendations for Legal Professionals

- ❖ Embrace technological proficiency - the difference between a junior lawyer soon to be struggling with finding viable employment and a CEO of one of those earlier mentioned legal start-ups may very well be a few Python and language technology courses. And if you prefer a more established corporate setting, understanding the mechanics of AI will put you at the forefront of candidates needed to aid big legal firms in their adaptation of AI.

- ❖ Pursue continuous learning - stay informed about any new AI-related regulations such as the AI Act currently under development in the EU. Just like the GDPR regulation caused many lawyers to successfully adopt it as their niche, AI-related regulations can be your domain.

- ❖ Adopt a collaborative mindset - the future of legal practice might very well lie, at least for the foreseeable future, in lawyers working alongside AI, harnessing its analytical power to release them from the most boring and bureaucratic aspects of their professions, while still providing added value in human insight and ethical considerations.

- ❖ Welcome the adaptation - humanity has always been defined by its adaptability. A mere half-century ago, the legal profession relied on typewriters for documentation. Since then, we have integrated computers, the internet, comprehensive legal databases, and sophisticated information retrieval systems into our daily practices, all within a lifetime of one legal professional. With that wisdom of hindsight, considering to now resist any of AI-related innovations seems almost senseless. Let us view this current AI shift as yet another fascinating chapter in the ongoing evolution of the legal profession.

- ❖ Championing ethical AI frameworks – lawyers and judges have traditionally stood at the intersection of societal change, ethical considerations, and the rule of law. In the face of rapid advancements in AI, legal professionals have a unique opportunity - and responsibility - to be integral voices in shaping the ethical contours of AI deployment. As interpreters and architects of statutes, norms, and rights, lawyers and judges can leverage their understanding of justice, equity, and the public interest to ensure AI systems are accountable, transparent, and respect fundamental human rights.

C. Future Research Directions

- ❖ Cross - Domain Replication Inclusive of Client’s Unique Circumstances

It would be valuable to replicate this study within different legal domains, for instance on a case related to a criminal law. Assessing how AI forms defence strategies or adapts to various legal terrains could enrich our understanding of its capabilities. Furthermore, incorporating a documented outline of a client's unique circumstances into model’s input could potentially further improve on this study results.

- ❖ AI’s Impact on Legal Education

It would be interesting to investigate how traditional legal writing is taught, and if tools such as GPT technology will change or challenge the expectations or standards in legal writing. We could examine the potential of AI tools to be used in legal education as either teaching aids or for practice exercises. Understanding both traditional legal writing education and the potential impact of AI would allow for a deeper exploration of how these two worlds might intersect.

❖ Client Perspective Study

This study examined the perspectives of legal professionals. Expanding the scope to include the perspectives of clients, who are the end recipients of the legal documents, but who are not legal professionals themselves would offer additional and valuable insights. It is a well-known fact that legal language often poses challenges for non-legal professionals due to its complexity and specialised terminology. AI-generated text, as observed in this study with Text B, has been praised for its clarity, potentially making it easier for clients to understand their own legal matters.

D. Final Thoughts

In the landscape where legal reasoning and argumentation are often viewed as uniquely human faculties, this research has provided an eye-opening perspective. AI, particularly GPT-4, has demonstrated a substantial ability not just to mimic but to craft effective legal arguments, outperforming human lawyers in the eyes of legal professionals themselves. The question 'Can AI make a case?' becomes a tangible reality. Yet, as the technology advances, the challenge becomes not simply recognising AI's capabilities but understanding how we, as legal practitioners, scholars, and responsible citizens, engage with this powerful tool. The future may not lie in AI vs. Lawyer but rather AI and Lawyer, working in partnership. The case, it seems, has only just been opened, and the next chapters promise to be as transformative as they are intriguing.

BIBLIOGRAPHY

1. Alkaissi, H., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing [Manuscript submitted for publication]. Cureus Journal of Medical Science. https://assets.cureus.com/uploads/editorial/pdf/138667/20230219-28928-6kcyip.pdf

2. Beioley, K., & Criddle, C. (2023, February 15). Allen & Overy introduces AI chatbot to lawyers in search of efficiencies. Financial Times. https://www.ft.com/content/baf68476-5b7e-4078-9b3e-ddfce710a6e2

3. CBS. (2019). Female majority in one-third of high-level occupations. Retrieved from https://www.cbs.nl/en-gb/news/2019/46/female-majority-in-one-third-of-high-level-occupations

4. CBS (2021). Origin: How many residents of the Netherlands were born abroad? Retrieved from https://www.cbs.nl/en-GB/visualisations/dashboard-population/origin

5. Clifton, J., & others. (2020). When machines think for us: The consequences for work and place. Cambridge Journal of Regions, Economy and Society, 13(1), 3–23. https://doi.org/10.1093/cjres/rsaa004

6. De Jong, P. O., & Herweijer, M. (2004). De ontwikkeling van het aantal wetten, AMvB's en ministeriële regelingen in Nederland. Overheid.nl - Alle regels tellen. Retrieved from https://zoek.officielebekendmakingen.nl/kst-29279-17-b5.pdf

7. Dekker-Abdulaziz, H. [@Hind_D66]. (2023). Parlementaire historie, zojuist een unieke motie ingediend! De ingediende motie is door ChatGPT gemaakt. Bij mijn weten de eerste keer dat er een motie is ingediend die gemaakt is met behulp van kunstmatige intelligentie. AI is ook een kans! [Tweet]. https://twitter.com/Hind_D66/status/1620753473343197186?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1620753473343197186%7Ctwgr%5E2766934fa7edc9ec668e832200cc24508261b718%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.geenstijl.nl%2F5169007%2Fhistorisch-d66-dient-door-ai-geschreven-motie-in%2F

8. Dutch Law. (2023). Lawyer in the Netherlands. https://dutch-law.com/lawyer-netherlands.html

9. Greene, J. (2022). Will ChatGPT make lawyers obsolete? (Hint: be afraid). Reuters. https://www.reuters.com/legal/transactional/will-chatgpt-make-lawyers-obsolete-hint-be-afraid-2022-12-09/

10. Helsten, J. L. (2019). Job Aid or Job Slayed? The Perceived Impact of Artificial Intelligence on Medical and Legal Work. Dissertation, Georgia State University. https://doi.org/10.57709/14336586

11. Hinkley, E. (2023). Miscon de Reya is hiring an 'engineer' to explore how its lawyers can use ChatGPT. Legal Cheek. https://www.legalcheek.com/2023/02/mishcon-de-reya-is-hiring-an-engineer-to-explore-how-its-lawyers-can-use-chatgpt

12. Katz, D. M., Bommarito, M. J. II, & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE, 12(4), e0174698. https://doi.org/10.1371/journal.pone.0174698

13. Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2023). GPT-4 Passes the Bar Exam. Illinois Tech - Chicago Kent College of Law; Bucerius Center for Legal Technology & Data Science; Stanford CodeX - The Center for Legal Informatics; 273 Ventures.

14. LegalFly. (n.d.). About. Retrieved from https://www.legalfly.ai/about

15. Macey-Dare, R. (2023). How ChatGPT and Generative AI Systems will Revolutionize Legal Services and the Legal Profession. St Cross College - University of Oxford; Middle Temple; Minerva Chambers. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4366749

16. Nolan, P. (2022). Artificial intelligence in medicine: How do we determine legal liability when things go wrong? (Thesis, Macquarie University). https://doi.org/10.25949/22138298.v1

17. OpenAI. (2023). GPT-4 Technical Report Retrieved from https://arxiv.org/abs/2303.08774

18. Ravenscroft, J., et al. (2021). CD2CR: Co-reference Resolution Across Documents and Domains. Computer Science > Computation and Language Cornell University. https://arxiv.org/abs/2101.12637v1

19. Robinson, J. (2014). Likert Scale. Michalos, A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_1654

20. Savelka, J., Ashley, K. D., Gray, M. A., Westermann, H., & Xu, H. (n.d.). Explaining Legal Concepts with Augmented Large Language Models (GPT-4). Computer Science Department, Carnegie Mellon University; Intelligent Systems Program, University of Pittsburgh; Cyberjustice Laboratory, Faculté de droit, Université de Montréal.

21. Schwarcz, D. B., & Choi, J. H. (2023). AI Tools for Lawyers: A Practical Guide [Manuscript submitted for publication]. Minnesota Law Review Headnotes. https://ssrn.com/abstract=4404017

22. Trustpilot (2023). Reviews of Juridisch Loket. Retrieved from https://nl.trustpilot.com/review/www.juridischloket.nl

23. Xiao, C., Hu, X., Liu, Z., Tu, C., & Sun, M. (2021). Lawformer: A pre-trained language model for Chinese legal long documents. Artificial Intelligence Open, https://doi.org/10.1016/j.aiopen.2021.06.003.

APPENDICES

1. Survey Form

Legal writing research

Thank you for participating in this exercise. You will be presented with a summary of a real legal case involving Employee_X and Company_X. Following the summary, you will find two legal letters (**Text A** and **Text B**) written in support of Employee_X's position. Please read the **Case Summary** and both letters with care and attention and then respond to a series of questions regarding the persuasiveness, clarity, and overall effectiveness of the arguments presented in each letter.

But first, please answer a few questions about you.

Switch account

Not shared

* Indicates required question

How old are you? *

- 20-30
- 31-40
- 41-50
- 51-60
- 61-70
- 71-80
- 80+

What is your gender? Pick one of the boxes below: *

- ○ Female
- ○ Male
- ○ Prefer not to say

What is/was your profession? Pick one of the boxes below: *

- ○ Lawyer
- ○ Paralegal
- ○ Judge
- ○ Legal Researcher
- ○ Law Professor
- ○ Legal Assistant
- ○ Other: ____________________

Case summary:

The client (Employee_X) is represented in a dispute regarding their employment contracts. They had two fixed-term contracts, followed by a third contract with a two-month extension. The client had their yearly bonuses always paid out and no registered performance issues until the contract conflict arose. The client argues that they should have an indefinite contract based on Dutch law, because they continued to work past that 2 month extension. The client is also involved in a dispute regarding the protection of confidential information. They have forwarded emails to their lawyer, but no sensitive data was shared.

Read both text A and text B. Then answer questions below it.

Text A:

TEXT A

Dear Colleague,

Thank you for your e-mail.

As stated before, my client always complied with the obligations set out in section 15 of her employment agreement. Please note that your client explicitly instructed my client that "confidential information should only be provided to or discussed (...) with third parties who have a legally binding obligation to keep the information confidential and a clear business need to receive the information." As I explained in our telephone conversation of last Wednesday, my client only forwarded those e-mails to herself at the request of her (former) lawyer (who has a legally binding obligation the keep the information confidential and instructed her to send those e-mails), so she would be able to prove in court that she was still actively working. Furthermore, my client only shared those e-mails with her (former) lawyer in a form of screenshot, thus without showing any sensitive data. Even if these screenshots would be considered to be confidential information – which my client denies – she was permitted to share those with her (former) lawyer based on your client's explicit instructions. My client therefore maintains that she has always complied with her confidentiality clause and is not liable to pay your client any damages as a result thereof. My client is very disappointed her ethics are being questioned and would be happy to prove the above in court to clear her name.

I have studied the judgement to which you refer in your e-mail. I have come to the conclusion that the facts of the aforementioned case differ on such crucial points that this judgement cannot serve as a basis for the case of my client. For example, in the case to which you refer, the employee had already signed a settlement agreement in which explicit confidentiality agreements were made, the employee in that case had no interest in forwarding the e-mails to prove - at the request of a lawyer - that he was still performing work and the e-mails he forwarded concerned quotations, price lists and price agreements. I therefore do not see the circumstance that a judge in the aforementioned case has decided to dissolve the employment contract as normative for my client's case, in which an entirely different set of facts and different contractual arrangements play a role.

I have had the opportunity to discuss your proposal with my client. First and foremost, please note that the proposal I made to you in our telephone conversation of last Wednesday is not the proposal you describe in your e-mail of 3 September 2020. Although we discussed the possibility of supplementing the unemployment benefits to match the salary, the proposal I made to you entailed a severance payment of 8 month's salary, which is an amount of € 40.840,- (gross). The proposal you have made therefore seems to be based on a misunderstanding.

Although my client is confident she will be successful in court in the claim for reinstatement because of her indefinite term employment agreement and in her defence to the accused breach of confidentiality, in order to end further discussions, I have found my client willing to agree to your proposal, provided that the following points are amended:

- *Increase of the severance payment to € 25.525,- (gross), which is the equivalent of five month's salary;*
- *Additional: € 5.015,- (gross) as a penalty for the absence of notification at the end of May 2020 (Section 7:668(3) DCC);*
- *Additional: compensation for costs of legal assistance, with a maximum of € 2.500,- excl. office costs and VAT;*

In view of the interests and legal position of my client, this is a very reasonable proposal. My client has made it very clear to me that she cannot agree to less. Please let me know in writing no later than ***Thursday 10 September at 15:00*** *whether your client can agree to the above proposal. If so, I will share a draft settlement agreement for your client to review. If your client does not agree with this proposal, I will initiate preliminary relief proceedings on behalf of my client, in which she will claim reinstatement. In that case, please inform me as soon as possible of the dates you and your client are unable to attend in court for the next eight weeks, so that they can be taken into account when requesting a date for a hearing.*

I await your timely response with interest and, in the meantime, I reserve all rights and defences on behalf of Employee_X.

Kind regards,
Employee_X lawyer

Text B:

TEXT B

Dear Colleague,

I hope this message finds you well. We appreciate your engagement in this matter and your efforts to seek a resolution. After careful consideration and consultation with Employee_X, we have prepared an alternative proposal that we believe better reflects the merits of our client's position under Dutch law, while still seeking a mutually agreeable settlement.

Firstly, I would like to address the issue of confidentiality. As stated previously, Employee_X has always adhered to the obligations set forth in Section 15 of their employment agreement. It is important to emphasize that any emails shared with their (former) lawyer were done so at the lawyer's request and without divulging any sensitive data. This exchange was carried out with a clear understanding of the necessity to maintain confidentiality while allowing Employee_X to demonstrate their active employment status. In light of these circumstances, Employee_X maintains that they have not breached any confidentiality clauses and should not be held liable for damages as a result.

In addition to the confidentiality matter, it is important to note that Employee_X has consistently received their yearly bonuses and has had no performance issues registered until the contract conflict arose. This further supports the notion that our client has been a valuable asset to your client's company, which should be taken into consideration when assessing the settlement.

Turning to the matter of employment contracts, we maintain that Employee_X should be granted an indefinite term employment agreement under Dutch law, given their continued work past the two-month extension period. As such, our client is confident in their legal standing should this matter proceed to court.

However, in the interest of resolving this dispute amicably, we would like to present the following amended proposal:

Severance payment increased to € 38,000 (gross), which is the equivalent of eight month's salary;
Additional compensation of € 5,015 (gross) as a penalty for the absence of notification at the end of May 2020 (Section 7:668(3) DCC);
Reimbursement for the costs of legal assistance, up to a maximum of € 3,500, excluding office costs and VAT.
We believe this proposal to be fair and reasonable given Employee_X's legal position, interests, and their commitment to your client's company. Our client has communicated that this is the minimum acceptable offer to reach a settlement. We kindly request that you provide a written response no later than Thursday, 10 September at 15:00, indicating whether your client accepts this proposal. Upon receiving your client's agreement, we will promptly share a draft settlement for review.

In the event that your client does not accept our proposal, Employee_X is prepared to pursue preliminary relief proceedings to claim reinstatement. If this becomes necessary, we kindly request that you provide us with any dates over the next eight weeks during which you and your client are unavailable for a court hearing.

We sincerely hope to reach a satisfactory resolution for both parties and appreciate your attention to this matter. In the meantime, we reserve all rights and defenses on behalf of Employee_X.

Kind regards,
Employee_X lawyer

1. **Overall Persuasiveness**: On a scale of **1** to **10**, rate the overall persuasiveness of Text A and Text B, where 1 is not persuasive at all and 10 is extremely persuasive. *

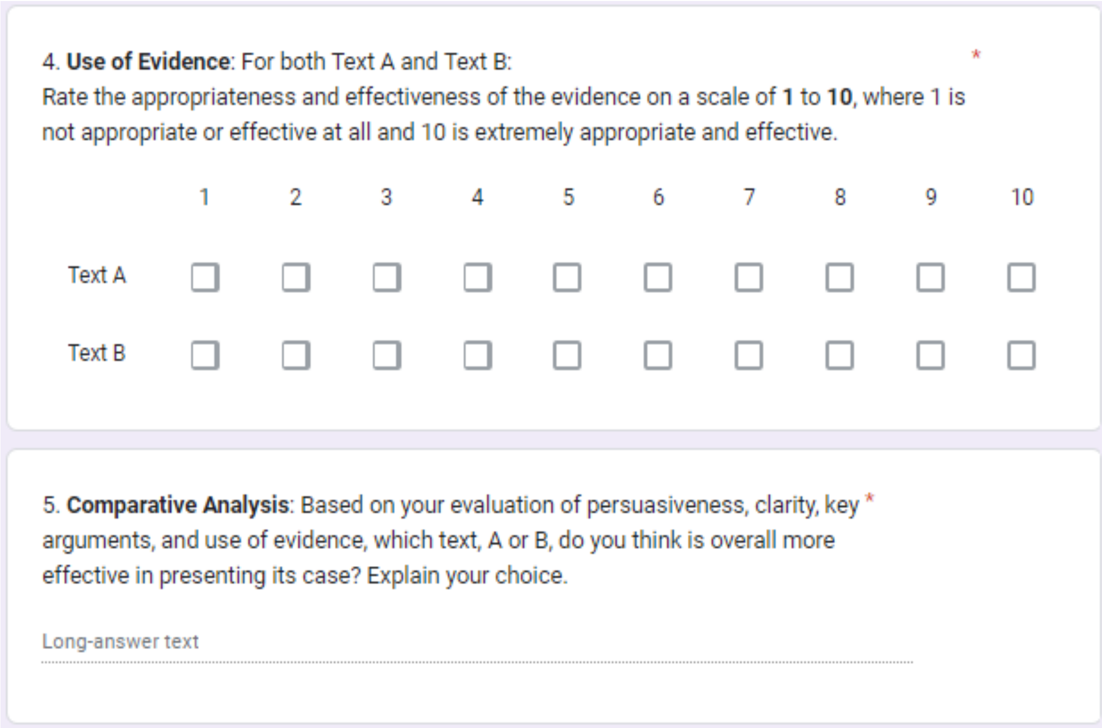
	1	2	3	4	5	6	7	8	9	10
Text A	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Text B	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

2. **Clarity and Coherence**: On a scale of **1** to **10**, rate the clarity and coherence of Text A and Text B, where 1 is not clear or coherent at all and 10 is extremely clear and coherent. *

	1	2	3	4	5	6	7	8	9	10
Text A	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Text B	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

3. **Strength of Key Arguments**: On a scale of **1** to **10**, rate the effectiveness of arguments in Text A and Text B, where 1 is not effective at all and 10 is extremely effective. *

	1	2	3	4	5	6	7	8	9	10
Text A	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Text B	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐


4. **Use of Evidence**: For both Text A and Text B: *

Rate the appropriateness and effectiveness of the evidence on a scale of **1** to **10**, where 1 is not appropriate or effective at all and 10 is extremely appropriate and effective.

	1	2	3	4	5	6	7	8	9	10
Text A	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Text B	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

5. **Comparative Analysis**: Based on your evaluation of persuasiveness, clarity, key * arguments, and use of evidence, which text, A or B, do you think is overall more effective in presenting its case? Explain your choice.

Long-answer text

2. Python code

```
#Installations of packages

!pip install openai;

!pip install pyPDF2;

!pip install tiktoken;


#Imports of libraries

import openai;

import PyPDF2;

import tiktoken; #count token


#

#User specific Variables

#

openai.organization = "$YOUR_ORG_HERE"
```

```
openai.api_key = "$YOUR_KEY_HERE"


#
#Thesis variables
#
API_MODEL = "gpt-4"
API_encoding = tiktoken.get_encoding("cl100k_base")
API_PROMPT_SYSTEM = "You are a non-biased Dutch legal counselor"
API_PROMPT_TEMP = 0 # Temperature value to 0, you will always see
the same response (most likely response). [0-1]
API_limit = 6500
API_max_tokens = 8000
API_DRYRUN = 0


#
#Functions
#
#Function to count tokens using tiktoken
def num_tokens_from_string(string: str, encoding_name: str) -> int:
    """Returns the number of tokens in a text string."""
    encoding = tiktoken.get_encoding(encoding_name)
    num_tokens = len(encoding.encode(string))
    return num_tokens


def calculate_total_tokens():
  All_Tokens = 0
```

```
    #Count the total amount of tokens to compreess
    for filename, description in File_Descriptions.items():


        #NLP Intro to the text
        current_document = "This text section contains: " + description + " "


        # Opening the files
        pdf_file = open(filename, "rb")
        pdf_reader = PyPDF2.PdfReader(pdf_file)


        # Count the amount of pages
        page_num_max=len(pdf_reader.pages)
        for page_num in range(page_num_max):
            page_text = pdf_reader.pages[page_num].extract_text().lower()
            current_document += page_text


        All_Tokens +=
num_tokens_from_string(current_document,'cl100k_base')


    return(All_Tokens)




def stage1_compression(buffer_document:str, buffer_Tokens: int,
text_compressed_total: str, tokens_compressed_total:int):
        #processing
        print (f'ChatGPT API - Called with Tokens: {buffer_Tokens}')
```

```
    #Clearing variables for next run

    text_compressed = ""

    tokens_compressed =0


    #Compression

    if ( API_DRYRUN != 1):

        response = openai.ChatCompletion.create(

            model=API_MODEL,

            messages=[

                    {"role": "system", "content":
f"{API_PROMPT_SYSTEM}"},

                    {"role": "user", "content": f"{API_PROMPT_USER}
TEXT: \"\"\"{buffer_document}\"\"\""},

            ],

            max_tokens=API_max_tokens-buffer_Tokens,    # Max=4097 -
input length

            temperature=API_PROMPT_TEMP        # Temperature value
to 0, you will always see the same response (most likely response).

            )

        text_compressed = response["choices"][0]["message"]["content"]


    #Statistics

    tokens_compressed =
num_tokens_from_string(text_compressed,'cl100k_base')


    print(f'From: {buffer_Tokens} to New size: {tokens_compressed}')


    #What would the output look like
```

```
print(f'\n==========================================\n{text_
compressed}\n-------------------------------\n')


        #Storing compressed text and total of compressed tokens

        text_compressed_total += text_compressed

        tokens_compressed_total += tokens_compressed

        return(text_compressed_total, tokens_compressed_total)



#

# (1) NLP - Manual Executed task

#



File_Descriptions = {

"20200827 Letter from Employee_X lawyer to Company_X_HR
ANON.pdf"          : "Letter from Employee_X lawyer to Company_X_HR
dated 20200827",

"Company_X lawyer email to Employee_X lawyer 20200831
ANON.pdf"              : "Company_X lawyer email to Employee_X lawyer
dated 20200831",

"Company_X_Response 20200803
ANON.pdf"                                                      :
"Company_X_Response to Employee_X dated 20200803",

"Employee_X - timeline of events
report_ANON.pdf"                                    : "timeline of events report
written by Employee_X",

"Employee_X lawyer email to Company_X lawyer 20200907
ANON.pdf"              : "Employee_X lawyer email to Company_X lawyer
dated 20200907",

"Registered letter from Employee_X to Company_X_HR 20200630
ANON.pdf" : "Registered letter from Employee_X to Company_X_HR
dated 20200630",
```

```
"Contract 2 for Employee_X prepared by Company_X ANON.pdf"          : "Contract 2 for Employee_X prepared by Company_X dated 20190614",

"Company_X lawyer email to Employee_X lawyer 20200910 ANON.pdf"          : "Company_X lawyer email to Employee_X lawyer dated 20200910",

"Contract 1 for Employee_X prepared by Company_X ANON.pdf"          : "Contract 1 for Employee_X prepared by Company_X dated 20180620"

}



# The following documents where not send to GPT

# Text A          - "Emplyee_X lawyer email to Company_X lawyer 20200915 ANON.pdf"          : "Emplyee_X lawyer email to Company_X lawyer dated 20200915"

# Outcome Text A    - "Settlement Agreement drafted by Company_X lawyer 20200916 ANON.pdf"    : "Settlement Agreement drafted by Company_X lawyer dated 20200916",




#

# (2) Prompt reduction - step1 individuals

#



API_PROMPT_USER="Your task is to compress the following text in a way that fits in a tweet (ideally) and such that you (GPT-4) \

can reconstruct the intention of the human who wrote text as close as possible to the original intention. \

This is for yourself. It does not need to be human readable or understandable. Abuse of language mixing, \

abbreviations, symbols (unicode and emoji), or any other encodings or internal representations is all permissible, \
```

```
as long as it, if pasted in a new inference cycle, will yield near-identical
results as the original"


document = ""

buffer_document = ""


Total_Tokens = 0

Count_Tokens = 0

Tokens = 0

buffer_Tokens = 0


text_compressed_total = ""

tokens_compressed_total = 0


# Calculate the total amount of tokens used

All_Tokens = calculate_total_tokens()


#Compression Step1 - Compress individuals

print( f'We need to compress the following total amount of tokens:
{All_Tokens}\n--------------------------------\n')


# Reading all available files

for filename, description in File_Descriptions.items():


    #NLP Intro to the text

    #current_document = "This text section contains: " + description + " "

    current_document = ""
```

```
    print(f'\tProcessing file {filename} ', end="")


    # Opening the files
    pdf_file = open(filename, "rb")
    pdf_reader = PyPDF2.PdfReader(pdf_file)
    # Count the amount of pages
    page_num_max=len(pdf_reader.pages)


    # Read all text of the pages and form a plaintext documents
    for page_num in range(page_num_max):
        page_text = pdf_reader.pages[page_num].extract_text().lower()
        current_document += page_text


    #Current size of the full document in tokens
    Count_Tokens =
num_tokens_from_string(current_document,'cl100k_base')
    print( f'({Count_Tokens} Tokens)')


    #Status of buffer to feed to ChatGPT API
    print(f'\t\tBuffersize: {buffer_Tokens}')


    #Create a buffer of less than 6500 tokens
    if ( buffer_Tokens + Count_Tokens < API_limit ): #cases: count > 6500,
count = 6500, count = 0-6499


        #appending to form buffer text
```

```
        buffer_document += current_document

        buffer_Tokens += Count_Tokens

        continue



    else:

        #processing with ChatGPT API

        text_compressed_total, tokens_compressed_total =
stage1_compression(buffer_document, buffer_Tokens,
text_compressed_total, tokens_compressed_total )



        #reset buffer

        buffer_Tokens = Count_Tokens

        buffer_document = current_document



#Last buffer empty run with ChatGPT API in case there is still buffer
remaining

if ( buffer_Tokens > 0):

    text_compressed_total, tokens_compressed_total =
stage1_compression(buffer_document, buffer_Tokens,
text_compressed_total, tokens_compressed_total )



# Final statistics of Compression stage 1

print(f'\n==================================================
======================\n')

print( f'Token count from: {All_Tokens} to:
{tokens_compressed_total}\nTextcompressed:\n {text_compressed_total}')




#
```

```
# (3) Generating Case Summary based on the pages fitting into Chatgpt
memory

#


Case_summary_inputtext = text_compressed_total

Case_summary_tokens = tokens_compressed_total

Case_summary_finaltext = ""


if ( Case_summary_tokens <= API_limit ):

   API_PROMPT_USER = "Your task is to Write an summary\n"


   #Running the case summary

   if ( API_DRYRUN != 1):

      response = openai.ChatCompletion.create(

      model=API_MODEL,

      messages=[

                  {"role": "system", "content":
f"{API_PROMPT_SYSTEM}"},

                  {"role": "user", "content": f"{API_PROMPT_USER}
TEXT: \"\"\"{Case_summary_inputtext}\"\"\""},

      ],

      max_tokens=API_max_tokens-Case_summary_tokens,    #
Max=4097 - input length

      temperature=API_PROMPT_TEMP         # Temperature value to 0,
you will always see the same response (most likely response).

   )


   Case_summary_finaltext =
response["choices"][0]["message"]["content"]
```

```
    print( f'Final Summary
text:\n==================================================\n{Ca
se_summary_finaltext}')

else:

    print( f'Case summary to long for current ChatGPT api
({Case_summary_tokens} / {API_limit} limit)')
```