

Executive Summary

The judgment is a key to understand the legal system and is the final output of the judicial system's handling of a case. It is not only for the parties to see, but also open to the whole society which shows how a specific case was handled and suggests how similar cases will be handled. Legal judgments in the UK are usually lengthy and complex documents, full of technical and intricate legal arguments.

Typically, cases heard by the UK Supreme Court (UKSC) involve a group of justices deliberating on the issue. The number of justices on the panel is usually odd to guarantee a definite majority result.ⁱ Following the deliberations, the justices engage in discussions and debates until they come to a consensus then one of them or the panel collectively draft the judgment. This process ensures that the final decision is reasonable, equitable, and represents the unified attitude of the UKSC.

Despite their significance, it is still difficult to obtain concise summaries of these rulings, which make effective legal study and understanding not that easy. Currently, all UK case judgments are manually summarized, with no available tool providing immediate access to summaries after verdict or offering customized summaries. This lack makes it more difficult to read through long legal documents and obtain important legal information in a timely manner.

This shortcoming highlights the urgent need for an innovative solution that automates the generation of legal document summaries based on user preferences and extract mentioned precedents cited in them. Our proposed response to address this need is the development of a cutting-edge AI-powered platform that leverages natural language processing and machine learning techniques to automatically generate case summaries and find relevant precedents.

Our planned next steps involve carrying out extensive research to improve the algorithms and methodologies for automated summarization and precedent extraction. After that, we will prototype the platform and test it to validate its effectiveness and accuracy, make sure it works as intended. Funding and collaborations are important for supporting the development, testing, and deployment stages of the platform. Additionally, we will focus on passing the platform through targeted marketing efforts and cooperating with legal institutions to guarantee broad acceptance and found if there's any improvement need. Through these actions, we aim to revolutionize legal research practices in the UK, advanced efficiency, accessibility, and comprehension in the legal field.

Evaluation of the Potential Opportunity

As the number of legal cases increases and documents become longer as social development issues become more complex, the need for efficient and high-quality access to case summaries becomes critical. From Figure 1 generated by LexisNexis, we found that there's an appetite to use generative AI for drafting documents (56%) and document analysis (44%). This means there's the chance and need to have automation or AI to help us in this field.

Most Potential for Generative AI in Legal					
	TOTAL	US	CANADA	FRANCE	UK
Researching matters	65%	59%	61%	77%	66%
Drafting documents	56%	53%	57%	55%	59%
Document analysis	44%	40%	40%	52%	47%
Writing emails	35%	30%	45%	34%	32%
Conducting due diligence	27%	27%	26%	18%	32%
Understanding new legal concepts	23%	19%	34%	24%	20%
Developing litigation strategies	12%	11%	12%	18%	9%
Other	4%	5%	3%	5%	4%
None / See no potential	5%	0%	10%	5%	7%

Figure 1: Forecast Use of Generative AI for Legalⁱⁱ

Currently, there are no specialized automated summarization tools used for legal texts in the UK like the one shown on the National Archive website. Although some techs or platform can complete the summarizing task, there are still needs in the market for our research to carried on. A quick look at the most recent judgments on the UKSC webpage indicates that their length can be in range from 10 to 70 pages¹ and imaging the situation we have to study hundreds of them when we are looking for the

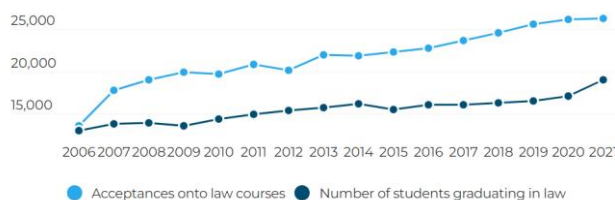
¹ <https://www.supremecourt.uk/decided-cases/index.html>

similar cases to support the views. Manual summarization is time-consuming and has delays, and users waiting for the website to publish or reading the entire judgment document themselves may cause delays in accessing critical legal information. The inefficiency and lack of accessibility in summarizing legal judgments in UK poses a significant challenge for public who rely on comprehensive and easily understandable summaries for various purposes, including case analysis, legal research and decision-making.

The opportunity lies in developing an intelligent processing system for specialized legal document, such as the one proposed in our project. Although there are general summary systems exist for various types of documents, they are not trained focus on the legal case dataset which means they cannot capture features required for processing legal texts effectively and their results may lead to oversimplify or misinterpret complex legal judgments. Our concept offers a unique opportunity to address the specific needs and challenges of the legal sector, making use of advanced natural language processing and machine learning techniques to extract key information, identify relevant precedents, and provide accurate and comprehensive summaries.

An automatic case summary generator has been implemented in a few legal jurisdictions outside the UK through the implementation of multiple research projects like summaries for Canadian judgmentsⁱⁱⁱ and judgments decided in Indian High Courts^{iv}. However, the same legal system is not enough to apply the exist platform on another e.g. Canada and the UK are both common law countries as the structure of legal reports and terminologies are different. These examples have the extensive results by introducing ROGUE metrics^v and Conditional Random Fields as a graphical technique which could be our alternative approaches.

Our solution will benefit a diverse range of stakeholders because everyone may have contact with the law. According to the survey, the number of students graduating with first degrees in law from universities in England and Wales in 2021 was the highest on record at 18927 and is increasing year by year. Law students need to study numerous cases to understand legal principles and precedents, laying



the foundation for their future careers. Our algorithm allows them to obtain summaries that cover all key points while accessing relevant cases, saving time and enhancing their understanding of complex legal concepts.

Figure 2 number of students accepted onto and graduating from first law degrees in England and Wales^{vi}

Meanwhile, the daily work of lawyers and legal professionals includes regularly reviewing judgments in preparation for defending clients, drafting legal documents, and providing legal advice. A small fact triggers a new precedent which means getting a quick summary speeds up information acquisition, allowing them to get more time to focus on strategic analysis and debate, and better find past examples that support their arguments.

The public can also benefit from simplified legal information when they have questions trying to understand the legal process or their rights and responsibilities under the law. Our idea can customize the presentation of summary or legal information and enable individuals to understand the legal system in an easy-to-understand manner. When they need basic legal assistance, they don't need to go offline for consultation with high costs and travel time.

Manual summaries of legal decisions may be subject to human error or inaccuracies, and different summarizers may interpret and prioritize information differently. Some views shared by a minority may not be revealed, and some cases contrary to society's default views are difficult to support. The automated summarization process can provide standardized and consistent summaries based on predefined criteria with further granularity on user-defined requirements, thereby increasing accuracy and helping to make more informed decisions and achieve fair outcomes in legal proceedings.

Value Proposition

We plan to develop an AI platform in the legal field that automatically generates legal document summaries customized to user specifications. The platform will use NLP and ML technologies to analyze user-uploaded PDFs or link of legal documents and extract key information, including relevant precedents, case details, regulations and connections between judges' views, and to a certain extent achieve visualization.

The platform plans to achieve several key goals:

1. By leveraging ML and NLP, we aim to generate concise and accurate legal document summaries based on user-defined preferences and requirements. This significantly reduces the time and effort required for manual summarization, increasing efficiency and productivity for legal professionals, researchers and students.
2. The algorithm automatically identifies and extracts precedents mentioned in the provided legal documents, allowing users to quickly identify and analyse them, facilitate comprehensive legal research and analysis, easily jump to mentioned details and reference standards.
3. Provides visualizing viewpoint tools, allowing users to gain deeper insights and understanding of legal documents through interactive visual representations and know the logic between various viewpoints.

The SUM system created by Hachey and Grover^{vii} which used supervised ML to identify the rhetorical label of a sentence in a judgment which categorising legal texts into 7 categories. Despite their contribution in the area being one of the best, the accuracy of their classifier is not high enough for practical use. We will try to use more advanced algorithms such as AdaBoost and LSTM, trying to improve the accuracy of the model and training on large datasets to improve effectiveness.

Valvoda et al. introduce Agreement Statements and Majority Opinion (ASMO)^{viii} which was developed to automatically identify two elements from UKHL cases. These key factors are claimed to have a significant impact in explaining the flow of a court speech and its conclusion. We will try to visualize the result in a graph to understand and shape the direction of our research which enables the platform to understand the structure and content of legal documents.

SUMmarisation with Majority Opinion (SUMO) introduced by Amy^{ix} abstractly and extractively summarising cases. By applying Conditional Random Fields (CRFs) and a bespoke summary generator on enriched data set, a pipeline capable of labelling, classifying, and extracting the most summary worthy sentences by using the sentence ranking scores and rhetorical role to generate the summaries was developed. Based on this, we will try to use some generation technologies to paraphrase the original text or generating new sentences instead of picking them up.

The platform uses machine learning models to automatically identify and extract relevant precedents mentioned in legal judgments. This process involves analyzing the text to identify references to prior cases, statutes, or legal principles that influenced court decisions. Users can customize the summaries generated by the platform according to their preferences and requirements. In addition, the platform provides visualization tools that present legal information in an interactive and easy-to-understand format, facilitating understanding and analysis. The further development of NLP and other technologies will make machine analysis closer to manual results and achieve better results. These technologies provide the basis for our implementation of this functionality, handling the complexities of legal language and reasoning more accurately.

Impact Plan

The entire development process will use the spiral model. In the first stage, we will do wide research, learn and reproduce the existing auto summarizing codes and the related models mentioned above, and explore ways to improve their research results. As their idea and performances are several years ago, they may mentioned some parsers no longer be maintained, we need to analyse the effectiveness of

them and find out the way to replace without any negative effect. Improve algorithms for automatic summarization and precedent extraction and explore the effects of different approaches using the latest appropriate ML and NLP techniques. We will connect the pipeline to the new National Archives data source where we can get the xml file as the largest dataset for all the cases. This phase may take six months, working with CS professions to implement the algorithms and methodologies, and try to achieve our visualization capabilities.

The second stage is when we complete the development of the entire system, we need to conduct a series of testing and validation, including black box testing and performance testing, to prove that our code efficiently and accurately completes our established goals, such as linking to National Archive data can be automatically redirected by obtaining access rights documents. We will use the manual description as the evaluation evidence to see if our methods achieve the satisfied results and suitable for all general cases. We will invite lawyers and law students to test, obtain valuable feedback and discover existing problems, such as whether the design meets the actual purpose of users.

When the test is completed, the automated summarization system will be deployed to provide a free version to the society and collect users' usage data and error feedback in the early stages which we can use to do the next round of research and development to make the code more perfect. We will conduct a user study to measure the usefulness of our system to see if it is a suitable alternative to manually written abstracts for ICLR. Continuous improvements and updates will be made based on user feedback and emerging trends in legal research.

Measurable Impacts

During the testing phase, we evaluate the accuracy and reliability of our project implementation by comparing the summaries produced by our system with the evaluations of human summaries and legal experts on the National Archive website aiming to achieve high accuracy (>90%).

In the subsequent phase of providing free use, measure the number of users who access the automatic summary tool, and track user engagement indicators, such as duration and frequency of use and effective feedback on the tool, to determine whether our research direction has produced meaningful results for users. Influence and provide efficient ways.

Quantify the cost savings users realize by using automated summarization tools compared to traditional manual summarization methods or subscription-based legal research platforms. Invite different users to conduct experiments with controlled variables to see if different tools are used to complete the same tasks. Positive impact on completion results, measuring the time users save using automated summarization tools for legal research and case analysis. Aim to reduce the time required for case analysis by at least 50%.

Funding

First, we invited the previously mentioned author Amy to cooperate as a person who has already done research in this field and recruited three other PhDs in CS to perform one-year developments. At the same time, we will recruit a law PhD student to join our development team as a legal-related background support, provide effective suggestions and help for our process, and provide timely feedback. We need to budget £1,000 for recruitment costs in addition to their labor costs.

In addition, we need to pay about 100,000 pounds to purchase equipment, some paid data, software and rent servers to store user usage data and our daily maintenance and system backup.

There is also a reserve of £10,000 for daily expenses and the purchase of other daily necessities, set aside to deal with unexpected or emergency expenses.

We estimate the total to be around £200,000.

Bibliography

ⁱ Casciani, D. (2020) What is the UK Supreme Court?, BBC News. Available at: <https://www.bbc.co.uk/news/uk-49663001>

ⁱⁱ International Legal generative AI report. Available at: <https://www.lexisnexis.com/pdf/lexisplus/international-legal-generative-ai-report.pdf>

ⁱⁱⁱ A. Farzindar and G. Lapalme, "Letsum, an automatic legal text summarizing system," Legal knowledge and information systems, JURIX, pp. 11–18, 2004

^{iv} M. Saravanan, B. Ravindran, and S. Raman, "Improving legal document summarization using graphical models," Frontiers in Artificial Intelligence and Applications, vol. 152, p. 51, 2006.

^v C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in Text Summarization Branches Out. Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 74–81. [Online]. Available: <https://www.aclweb.org/anthology/W04-1013>

^{vi} International Legal generative AI report. Available at: <https://www.lexisnexis.com/pdf/lexisplus/international-legal-generative-ai-report.pdf>

^{vii} B. Hachey and C. Grover, "Extractive summarisation of legal texts," Artificial Intelligence and Law, vol. 14, no. 4, pp. 305–345, 12 2006.

^{viii} J. Valvoda, O. Ray, and K. Satoh, "Using agreement statements to identify majority opinion in ukhl case law," in Frontiers in Artificial Intelligence and Applications, M. Palmirani, Ed., vol. 313. Netherlands: IOS Press, 12 2018, pp. 141–150

^{ix} O. Ray, A. Conroy, and R. Imansyah, Summarisation with majority opinion., in JURIX, 2020, pp. 247–250.