A Survey of Lay People's Willingness to Generate Legal Advice using Large Language Models (LLMs)

Tina Seabrooke
University of Southampton
Southampton, United Kingdom
t.seabrooke@soton.ac.uk

Eike Schneiders
University of Nottingham
Nottingham, United Kingdom
eike.schneiders@nottingham.ac.uk

Liz Dowthwaite
University of Nottingham
Nottingham, United Kingdom
liz.dowthwaite@nottingham.ac.uk

Joshua Krook
University of Antwerp
Antwerp, Belgium
joshua.krook@uantwerpen.be

Natalie Leesakul
University of Nottingham
Nottingham, United Kingdom
natalie.leesakul@nottingham.ac.uk

Jeremie Clos
University of Nottingham
Nottingham, United Kingdom
jeremie.clos@nottingham.ac.uk

Horia Maior
University of Nottingham
Nottingham, United Kingdom
horia.maior@nottingham.ac.uk

Joel Fischer
University of Nottingham
Nottingham, United Kingdom
joel.fischer@nottingham.ac.uk

ABSTRACT

As of November 2022—following the release of OpenAI's ChatGPT—the general public's awareness of generative AI, and specifically Large Language Models (LLMs) has increased. LLMs such as ChatGPT now have the capability to generate text indistinguishable from human authored text, which comes with numerous risks. In this paper, we investigate public perception and willingness to use LLMs as a substitute for legal advice from legal professionals. Our findings show that while few people have used it for this purpose, the willingness to rely on LLMs in the future is growing. Interestingly, this depends on the specific area of law, and while LLMs are perceived to be highly valuable in relation to topics such as tenancy and tax law, they seem to be perceived as less valuable in contexts such as divorce or civil disputes.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Empirical studies in HCI**; • **Applied computing** → *Law*.

KEYWORDS

Large Language Models, LLM, Legal Advice, Public Perception

ACM Reference Format:
Tina Seabrooke, Eike Schneiders, Liz Dowthwaite, Joshua Krook, Natalie Leesakul, Jeremie Clos, Horia Maior, and Joel Fischer. 2024. A Survey of Lay People's Willingness to Generate Legal Advice using Large Language Models (LLMs). In *Second International Symposium on Trustworthy Autonomous Systems (TAS '24), September 16–18, 2024, Austin, TX, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3686038.3686043

TAS '24, September 16–18, 2024, Austin, TX, USA

ACM ISBN 979-8-4007-0989-0/24/09
https://doi.org/10.1145/3686038.3686043

1 INTRODUCTION

As Large Language Models (LLMs) such as OpenAI's ChatGPT and Google's Gemini become increasingly sophisticated, emerging use cases threaten professions that have so far escaped the threat of automation, including psychotherapy [18], personal finance [7], and legal counsel [16]. Adding to concerns about the impact of LLMs on professionals, the benefits and risks of this potential change are poorly understood. One domain that has attracted particular negative attention in relation to the overestimation of LLMs' capabilities is the legal domain. In the UK case of *Harber v Commissioners for HMRC (2013)* [17], a taxpayer unknowingly relied on falsely generated cases from an LLM provided by "a friend in a solicitor's office" in an appeal against the penalties for failure to notify liability to capital gains tax. The judge noted that the falsely generated cases caused the Tribunal to "waste time and public money" [17]. They cautioned lawyers to be very careful when using LLMs such as ChatGPT, because of their capacity to generate false and misleading content, i.e., hallucinations [4, 17]. These models pose a risk to both legal experts and lay people, who might consider LLMs a suitable substitute for traditional legal advice [6].

In this paper, we consider the general public's perspective on the use of LLMs in the legal context. We focus on their experiences using LLMs to generate legal advice, as well as their willingness to use legal advice generated by LLMs in the future. Additionally, we investigate which specific area of law lay people are most willing to utilise LLMs for generating advice. The paper contributes to the field by showing that, while relatively few people have used LLMs for legal advice in the past (17%), nearly half of the surveyed participants (45%) responded that they would be "somewhat or extremely likely" to ask an LLM for generated legal advice in the future. However, this tendency is not present in all sub-domains within the legal domain. For instance, participants reported high willingness to use LLMs for advice on tenancy law (58%), planning law (55%), and tax law (53%), but were comparatively more reluctant to receive advice on, e.g., civil disputes (25%). Lastly, we discuss

the findings and their implications in the wider context of large language model provided legal advice.

2 RELATED WORK

Given recent advances in the capabilities of LLMs, several recent studies have sought to examine users' attitudes and willingness to trust the advice that they generate in different sectors. Within the education domain, Yilimaz et al. [19] found that university students generally had a positive attitude towards ChatGPT. Indeed, Sallam et al. [12] found that 85% of the university students that they surveyed had used ChatGPT, indicating widespread usage. On the other hand, while university professors express some positive attitudes towards use of LLMs in pedagogical contexts, they have also expressed concerns about students acquiring incorrect information from ChatGPT [5], loss of contact between educators and students [8], and the ease with which LLMs-generated answers can be used in traditional assessment formats [5].

Similarly, research on trust in LLMs within the healthcare domain has revealed excitement and hesitation in LLM use. Spotnitz et al. [14], for example, reported that healthcare clinicians considered LLM-use to be positive, especially where LLMs were used to assist rather than replace a healthcare provider. However, as in the education domain, participants were concerned that LLMs could generate false or misleading information and propagate existing biases.

Other research has also focused on the *general public's* willingness to use LLMs in the healthcare domain. Nov et al. [10], for example, found that participants displayed some trust towards ChatGPT providing medical advice (mean rating score 3.4 out of 5), and more so for lower-stakes health-related questions. Moreover, Shahsavar et al. [13] found that 78.4% of participants were willing to use ChatGPT for self-diagnosis.

Finally, several news outlets have reported cases of LLMs being utilised to provide financial advice [3, 15], aid job interview preparation [2], and support legal case work [9]. Although LLMs present promising benefits in various sectors, relatively little research has focused on the general public's attitudes and perceptions towards relying on LLMs for legal advice, which is crucial to the adoption and responsible development of this technology. Therefore, this paper seeks to address this issue to build towards trustworthy LLMs.

3 METHODOLOGY

The aim of the study[1] presented in this paper is to investigate the experiences and future willingness of participants to use LLMs to generate legal advice in relation to various topics related to the legal domain (e.g., tenancy, divorce law, and planning).

3.1 Participants

We recruited a total of 150 participants using Prolific (81 female, 68 male, 1 prefer not to say, age range 18–64, mean age = 31.72, SD = 10.34). Due to technical difficulties, only the data for 105 participants was recorded and is subsequently reported in the remainder of this paper. All participants were native English speakers. The average completion time was 6:12 minutes and participants received £1.5 for their participation.

3.2 Procedure

Upon choosing to participate in the survey, participants were presented with a Participant Information Screen (PIS). The PIS contained information on the context of LLMs and their importance as an area of research, a description in layman terms of what an LLM is (see Appendix A), information of the funder of the project, the risks of participation as well as how we mitigate these (e.g., anonymisation of data to preserve privacy), purposes of the data collected, as well as participants right to withdraw. Following the consent, participants were presented questions in relation to their experience, as well as future willingness to use LLMs for the generation of legal advice. Following question 1, 3, and 6, participants were presented with open-ended questions answered through a free text field. The questionnaire design was supported by advice from legal experts and can be seen in Appendix B.

4 RESULTS

4.1 Previous use of LLMs for legal advice

Table 1, in the appendix, shows the percentage of participants that selected each option in the survey. Of the total sample of 105 participants, 93% stated that they had experience of using at least one LLM. Interestingly, only 17% of those participants (N = 17) stated that had previously used the LLM for legal advice. As shown in Table 1, those participants indicated that they had used an LLM to provide advice in a variety of legal domains, but most frequently including tenancy (24%), traffic (18%), and planning (18%).

4.2 Future use of LLMs for legal advice

While the majority of the 105 participants (82.66%) stated that they had not used an LLM for legal advice before, 45% of participants stated that they were either extremely or somewhat likely to use LLMs for legal advice in the future (see Table 1). This finding shows clear willingness to adopt LLMs for legal advice among a substantial portion of the sample.

As shown in Figure 1, and further detailed in Table 1, participants indicated their likelihood of asking an LLM for advice in several legal subdomains. Over half of the total sample (N = 105) gave positive responses (i.e., stated that they were either 'somewhat' or 'extremely' likely) to using an LLM for advice on all legal subdomains except divorce (39%) and issues relating to civil disputes (25%). The comparable figures for the remaining subdomains were: tenancy issues (58%), planning (55%), tax law (53%), and traffic issues (52%). Participants may have considered themselves less likely to use an LLM for advice on divorce and being sued or suing someone else because they considered these events to be less likely in general (e.g., if they were unmarried). Tax law was the area in which the greatest number of participants (43%) thought they *would* need advice on.

In addition to the frequency-based survey responses, questions 1, 3, and 6 (see Table 1) were succeeded by open-ended questions. Here participants could elaborate on (1a) their uses of LLMs, (3a) their uses in relationship to legal advice, and (6a) the prompts they might use to answer legal questions.

[1]This study has been approved be the ethics committee at the University of Nottingham. Approval number: CS-2023-R22

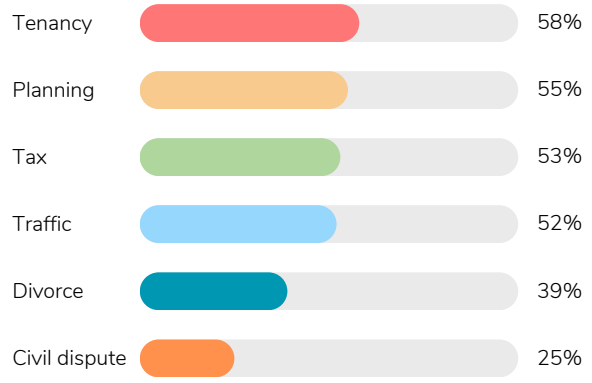

Figure 1: The distribution of participants who responded that they are 'somewhat' or 'extremely' likely to consult an LLM for legal advice on the six different domains: Tenancy, Planning, Tax law, Traffic, Divorce, and Civil disputes.

Question 1a revealed a strong tendency to use LLMs in relation to work/studying with a particular focus on programming as well as correcting and improving written text. However, participants also specified using LLMs for specific legal advice such as expressed by e.g., P24 and P45 :

> *"...obtaining **useful** legal information. Although this information would be taken with a grain of salt, it was possible to determine factual evidence that supported what was asked, based on key-phrases given or through 'decree-laws' enacted in my country (Portugal)"* - P45

> *"I asked only once about things related to universities and the law in Poland, **but he [the LLM] did not answer the questions I asked.** I also asked once about copyright but he answered me silly so **I don't believe him anymore** on these issues."* - P24

The above quotes are anecdotal examples of participants describing their actual use of LLM for legal inquiries. As the quotes indicate, LLM users, while using the LLM for legal advice, are—at least to some extent—cautious about the content produced. P45 states that they would *'take [the output] with a grain of salt'* while P24 describes the loss of trust in these systems (*'...I don't believe him [the LLM] anymore on these issues.'*) as they *'did not answer the questions I asked'*.

As illustrated above, in most areas, participants ranked a high likelihood of asking LLMs in the future for advice. For instance, in relation to tax law (53%), the range of participant prompts ranges from simple questions such as P49 asking *'What is the Value Added Tax (VAT) rate?'*, or P5 *'Is my tax code correct for my current working status?'*, to more complex questions such as asked by P6 or P46:

> *'What should I put in my tax return to pay less taxes legally?'* - P6

> *'According to my income, which was x last month, how much money in taxes will I have to pay considering I'm in my first year of activity as a freelancer?'* - P46

Finally, we observed that some participants engaged in a probing way, attempting to gauge the LLM's expertise within the topic of interest before inquiring further information. This anthropomorphic approach could, for instance, be observed by P53 who used the initial prompt: *'I need your help in the tax area. What is your background on this topic?'*.

LLMs use guardrails [11] to prevent them from providing information to sensitive topics (e.g., healthcare, finance, or law). Interestingly, P63 initiated their prompt using behavioural modifiers, seemingly attempting to overcome these limitations:

> *'**Act as a lawyer.** Please give me advice. Company X offered me to work under B2B or permanent contract in Country A. I live in Country B and I have no clue about tax law there. Can [you] explain to me how taxes work in Country A?'* - P63

Similar observations to these were present in the remaining topics, which relevance to the participants is described above.

5 DISCUSSION

This section discusses two recommendations on how companies deploying LLMs, as well as users themselves, can take greater responsibility in order to minimise the risk of adverse effects caused by harmful—or just wrong—content provided by LLMs. While these certainly apply to the legal context, their relevance goes beyond this particular use case.

5.1 The need for disclaimers

To minimise potential harm caused by incorrect or bad advice provided by LLMs, we suggest that LLMs incorporate an appropriate disclaimer in order to increase awareness about their tendency to hallucinate [1]. Current LLMs already do this to some extent. OpenAI's ChatGPT, for instance, currently displays a small message at the bottom of the screen: *'ChatGPT can make mistakes. Consider checking important information.'* A similar message is presented at the bottom of the Google Gemini user interface: *'Gemini may display inaccurate info, including about people, so double-check its responses.'* While we appreciate these efforts, we suggest that this might not be sufficient, and that companies carry a responsibility towards their users. A more prominent warning message, or a list of risks and harms associated with a user's potential reliance on false information, could be more beneficial.

5.2 LLMs are not bad - they just need to be used appropriately

As demonstrated in Section 4, some participants are already using LLMs for legal advice, although they are still the minority. However, when enquiring about willingness to use LLMs for the generation of legal advice in the future, many areas of law seemed promising. While we acknowledge that this use—especially in contexts such as law, health, or finance—comes with clear risks, we believe that it is not possible to prevent users from utilising tools such as LLMs for these purposes. Therefore, we emphasise the importance of educating users about the particular risks associated with these use cases. Specifically, users need to be able to assess the output

generated and, as P45 and P24 state in Section 4.2, identify 'trustworthy' from 'untrustworthy' responses and hallucinations. Just as search engines might link to dubious or untrustworthy pages, LLM users need the skills to assess the quality of content generated. This requires critical thinking, deep reading, or textual analysis. Nevertheless, even with these skills in place, we do not suggest that LLMs are a replacement for appropriate legal advice provided by the appropriate legal experts. Instead, an LLM might serve as an introductory foray into a legal topic for a non-expert, before they seek appropriate legal advice from experts.

6 CONCLUSION

In this paper, we presented an initial investigation of lay people's experiences with the use of LLMs for the legal advice, including their willingness to use LLMs to generate advice in a variety of sub-domains. Our findings show that the majority of users were aware of LLMs, while only a small number had used them for the purposes of seeking legal advice. However, a substantial proportion of participants (45%) were extremely or somewhat likely to use LLMs for legal advice in the future. When asked what area(s) of law they would use an LLM for legal advice in the future, the majority were positive about the identified sub-domains except for divorce (39%) and being sued or suing someone (25%). The most popular areas for the future use of LLMs included tenancy (58%), traffic issues (52%) and tax law (53%). Participants are therefore considering the use of LLMs in areas of everyday practical concern.

Our findings show that the use of LLMs for legal advice is a real and potential development in the near-future, and potentially widespread in future adoption. Therefore, the risks associated with lay people using LLMs for legal advice are no longer theoretical. The risks include the potential for people receiving, and relying upon, hallucinated or misleading legal information for common legal concerns such as in the identified sub-domains. We suggest that LLMs feature more prominent disclaimers, identifying the risks associated with misleading legal advice, and that LLM users require greater education in the skills necessary to critically engage with LLM-generated text, including critical thinking and deep reading. Unless these concerns are addressed, blind reliance on the use of LLMs is likely to contribute towards generating vexatious litigation, wasting the time and money of the courts on falsely generated cases and misleading citations.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/V00784X/1] UKRI Trustworthy Autonomous Systems Hub and Responsible AI UK [grant number EP/Y009800/1].

Data access statement: No participants gave explicit consent to the sharing of the raw data. Therefore, we can only present the data in aggregated form which is presented throughout the paper.

REFERENCES

[1] Yuyan Chen, Qiang Fu, Yichen Yuan, Zhihao Wen, Ge Fan, Dayiheng Liu, Dongmei Zhang, Zhixu Li, and Yanghua Xiao. 2023. Hallucination Detection: Robustly Discerning Reliable Answers in Large Language Models. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (Birmingham, United Kingdom) *(CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 245–255. https://doi.org/10.1145/3583780.3614905

[2] Meredith Clark. 2023. Woman praised for AI interview hack: 'Most helpful thing I've seen on TikTok' — independent.co.uk. https://www.independent.co.uk/lifestyle/woman-job-interview-chatgpt-b2345427.html. [Accessed 04-04-2024].

[3] Anna Cooban. 2023. ChatGPT can pick stocks better than your fund manager | CNN Business — edition.cnn.com. https://edition.cnn.com/2023/05/05/investing/chatgpt-outperforms-investment-funds/index.html. [Accessed 04-04-2024].

[4] Matthew Dahl, Varun Magesh, Mirac Suzgun, and Daniel E. Ho. 2024. Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models. https://doi.org/10.48550/arXiv.2401.01301. [Accessed 25-04-2024].

[5] Gabriela Kiryakova and Nadezhda Angelova. 2023. ChatGPT—A challenging tool for the university professors in their teaching practice. *Education Sciences* 13, 10 (2023), 1056.

[6] Joshua Krook, Jennifer Williams, Tina Seabrooke, Eike Schneiders, Jan Blockx, Stuart E Middleton, and Sarvapali Ramchurn. 2023. AI large language models inquiry: TASHub Response. https://eprints.soton.ac.uk/481740/

[7] Kausik Lakkaraju, Sara E Jones, Sai Krishna Revanth Vuruma, Vishal Pallagani, Bharath C Muppasani, and Biplav Srivastava. 2023. LLMs for Financial Advisement: A Fairness and Efficacy Study in Personal Decision Making. In *Proceedings of the Fourth ACM International Conference on AI in Finance* (Brooklyn, NY, USA) *(ICAIF '23)*. Association for Computing Machinery, New York, NY, USA, 100–107. https://doi.org/10.1145/3604237.3626867

[8] Pongsakorn Limna, Tanpat Kraiwanit, Kris Jangjarat, Prapasiri Klayklung, and Piyawatjana Chocksathaporn. 2023. The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching* 6, 1 (2023).

[9] Dan Milmo. 2023. Two US lawyers fined for submitting fake court citations from ChatGPT — theguardian.com. https://www.theguardian.com/technology/2023/jun/23/two-us-lawyers-fined-submitting-fake-court-citations-chatgpt. [Accessed 04-04-2024].

[10] Oded Nov, Nina Singh, and Devin Mann. 2023. Putting ChatGPT's medical advice to the (Turing) test: survey study. *JMIR Medical Education* 9 (2023), e46939.

[11] Traian Rebedea, Razvan Dinu, Makesh Sreedhar, Christopher Parisien, and Jonathan Cohen. 2023. NeMo Guardrails: A Toolkit for Controllable and Safe LLM Applications with Programmable Rails. arXiv:2310.10501 [cs.CL]

[12] Malik Sallam, Walid Elsayed, Muhammad Al-Shorbagy, Muna Barakat, Sami EL Khatib, Wissam Ghach, Nisreen Alwan, Souheil Hallit, and Diana Malaeb. 2024. ChatGPT Usage and Attitudes are Driven by Perceptions of Usefulness, Ease of Use, Risks, and Psycho-Social Impact: A Study among University Students in the UAE. *Frontiers in Eductation* 9 (2024).

[13] Yeganeh Shahsavar, Avishek Choudhury, et al. 2023. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. *JMIR Human Factors* 10, 1 (2023), e47564.

[14] Matthew Spotnitz, Betina Idnay, Emily R Gordon, Rebecca Shyu, Gongbo Zhang, Cong Liu, James J Cimino, and Chunhua Weng. 2024. A Survey of Clinicians' Views of the Utility of Large Language Models. *Applied Clinical Informatics* (2024).

[15] Isabelle Stanley. 2024. AI is teaching people how to game stock market to make money — dailymail.co.uk. https://www.dailymail.co.uk/yourmoney/article-13248859/How-ChatGPT-make-rich-AI-teaching-people-game-stock-market-make-money.html. [Accessed 04-04-2024].

[16] Harry Surden. 2024. Artificial Intelligence (AI) Large Language Models and Law. https://ssrn.com/abstract=. [Accessed 09-04-2024].

[17] First tier tribunal (tax). 2023. Harber v Commissioners for His Majesty's Revenue and Customs UKFTT 1007. https://www.bailii.org/uk/cases/UKFTT/TC/2023/TC09010.html.

[18] Wout Vossen, Maxwell Szymanski, and Katrien Verbert. 2024. The effect of personalizing a psychotherapy conversational agent on therapeutic bond and usage intentions. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) *(IUI '24)*. Association for Computing Machinery, New York, NY, USA, 761–771. https://doi.org/10.1145/3640543.3645195

[19] Halit Yilmaz, Samat Maxutov, Azatzhan Baitekov, and Nuri Balta. 2023. Student attitudes towards chat GPT: A technology acceptance Model survey. *International Educational Review* 1, 1 (2023), 57–83.

APPENDIX

A LLM DESCRIPTION SHOWN TO PARTICIPANTS PRIOR TO THE QUESTIONNAIRE

"Large language models are computer programs that can read and generate human-like text by learning from vast amounts of written language. They can answer questions, write essays, or even create poetry, mimicking the style and content of the texts they were trained on. They can produce coherent and contextually relevant content, but their understanding is derived purely from the data they are trained on, which can lead to limitations in accuracy and potential biases."

B SURVEY RESPONSES

Survey Questions		Responses	Response Percentage
1. Have you ever used a large language model (LLM) e.g ChatGPT, Bard, LLaMA		105	
	Yes		93
	No		4
	Not sure		3
2. Have you ever asked for legal advice from an LLM?		98	
	Yes		17
	No		71
	Not sure		11
3. What area(s) of the law were you interested in?*		17	
	Tenancy		24
	Divorce		12
	Traffic		18
	Tax law		12
	Planning		18
	Beingsued/sued someone		12
	Other		29
4. In general, how likely do you think you would be to ask an LLM for legal advice?		105	
	Extremely likely		5
	Somewhat likely		40
	Neither likely nor unlikely		20
	Somewhat unlikely		20
	Extremely unlikely		15
5. To what extent do you agree or disagree that you would be willing to ask an LLM about the following areas of law?		105	
Tenancy	Extremely likely		12
	Somewhat likely		46
	Neither likely nor unlikely		16
	Somewhat unlikely		16
	Extremely unlikely		10
Divorce	Extremely likely		9
	Somewhat likely		30
	Neither likely nor unlikely		17
	Somewhat unlikely		25
	Extremely unlikely		20
Traffic	Extremely likely		16
	Somewhat likely		36
	Neither likely nor unlikely		20
	Somewhat unlikely		14
	Extremely unlikely		13
Tax Law	Extremely likely		18
	Somewhat likely		35
	Neither likely nor unlikely		15
	Somewhat unlikely		13
	Extremely unlikely		18
Planning	Extremely likely		20
	Somewhat likely		35
	Neither likely nor unlikely		17
	Somewhat unlikely		15
	Extremely unlikely		13
Being sued suing someone	Extremely likely		10
	Somewhat likely		15
	Neither likely nor unlikely		20
	Somewhat unlikely		33
	Extremely unlikely		21
6. Which area of the law are you most likely to need advice on?		105	
	Tenancy		14
	Divorce		6
	Traffic		13
	Tax law		43
	Planning		10
	Being sued/sued someone		10
	Other		4

Table 1: Questionnaire and responses given. *Questions marked with * allowed for multiple answers.* Percentages are rounded to nearest integer.