Considerations when Using ChatGPT and Generative Artificial Intelligence Software Based on Large Language Models (Jan. 2024)

The Bar Council (UK)

© Crown copyright 2023. Licensed for re-use under the Open Government License v3.0.

The Bar Council

Considerations when using ChatGPT and generative artificial intelligence software based on large language models

Purpose:	To provide barristers with a summary of considerations if using ChatGPT or any other generative AI software based on large language models (LLMs).
Scope of application:	All barristers and chambers
Issued by:	The Information Technology Panel.
Issued on:	30 January 2024
Status and effect:	**Please see the notice at end of this document. This is not "guidance" for the purposes of the BSB Handbook I6.4.**

Introduction

1. In the rapidly evolving technological landscape, particularly with the evolution of Generative Artificial Intelligence (**Generative AI**) based on large language model (**LLM**) systems, like OpenAI's ChatGPT and Google's Bard, being used by legal professionals, the Bar Council has issued this guidance to assist barristers in understanding the technological basis and risks in the use of such generative LLM systems. Although this guidance may not be exhaustive, it provides the main considerations for the use of LLMs, in order for barristers to adhere to legal and ethical standards; safeguarding client confidentiality, and maintaining trust and confidence, privacy and compliance with applicable laws.

2. The purpose of this guidance is to provide a useful summary of considerations for barristers if they decide to use ChatGPT or any similar LLM software. It should also be noted that generative LLM technologies are developing rapidly and as the field of generative AI continues to evolve, with new models and advances being introduced

regularly, it is always good to understand the underlying model and acknowledge its limitations prior to using these technologies. It is important to note that the legal and regulatory landscape on the use of AI is subject to constant change, and therefore barristers will need to be vigilant and adapt accordingly.

What is large language model (LLM) software?

3. It is easier to begin by explaining what it is not. It is not a conventional research tool, it does not analyse the *content* of data and it does not think for itself. It is, rather, a very sophisticated version of the sort of predictive text systems that people are familiar with from email and chat apps on smart phones, in which the algorithm predicts what the next word is likely to be. LLMs use machine learning algorithms, first to be 'trained' on text and, based on that 'training' (which involves the application of *inter alia* mathematical formulae), to generate sequential text. These programmes are now sufficiently sophisticated that the text often appears as if it was written by a human being, or at least by a machine which thinks for itself.

4. LLMs have not been around long enough and have not been sufficiently tested for it to be clear what tasks they can or should be used for in legal practice. Some practitioners and judges have made positive comments about using them to arrange text. However, **it is important for barristers who choose to use LLMs to do so responsibly and think about what they are doing, by weighing the potential risks and challenges associated with such use in the light of their professional responsibilities.**

What is ChatGPT?

5. ChatGPT is an advanced LLM AI technology developed by OpenAI. It is based on GPT architecture, which stands for 'Generative Pre-Trained Transformer'. The latest iteration of ChatGPT at the time of this guidance is GPT-4. Transformer architecture uses mathematical matrices, supplemented by corrective procedures and technologies. The number of parameters used by GPT-4 is thought to be in the many billions.

6. In common with other LLMs (such as Google's Bard), ChatGPT is trained on huge amounts of data, which is processed through a neural network made up of multiple nodes and layers. These networks continually adjust the way they interpret and make sense of data based on a host of factors, including the results of previous trial and error.

7. Certain consequences inevitably follow from the nature of the technological process that is being carried out. LLM AI systems are not concerned with concepts like 'truth' or accuracy.

Key risks with LLMs

8. **Anthropomorphism:** The first key risk inherent in LLMs is that they are designed and marketed in such a way as to give the impression that the user is interacting with something that has human characteristics. One of the mechanisms by which this is sought to be achieved is by the use of anthropomorphic language to describe what is happening. Perhaps the most obvious example of this is the use, by OpenAI, of the word 'Chat' in the name of its LLM products (ChatGPT). As set out above, LLMs (at least at the current stage in their development) do not have human characteristics in any relevant sense.

9. **Hallucinations:** It has been said that LLMs are prone to "hallucinations", a term which is used to describe the phenomenon where the outputs generated by these LLMs may sound plausible but are either factually incorrect or unrelated to the given context.[1] However, whilst the use of this term is helpful for illustrative purposes it demonstrates the widespread tendency to anthropomorphise the technology. As we say, it is necessary when using LLMs to keep well in mind the actual technical process that is being carried out.

10. **Information disorder:** The ability for ChatGPT inadvertently to generate information disorder, including misinformation, is a serious issue of which to be aware. This is not only because of the volume of misinformation that is likely to be in the training data.

11. These points are illustrated in the affidavit that was filed by a New York lawyer to explain his conduct when he included six fictitious cases suggested by ChatGPT in his submissions.[2] Screenshots of what he had done were annexed to his affidavit:

> *"Is varghese a real case"*
>
> *"Yes, Varghese v. China Southern Airlines Co Ltd, 925 F. 3d 1339 (11th Cir. 2019) is a real case."*
>
> *"What is your source"*
>
> *"I apologize for the confusion earlier. Upon double-checking, I found that the case Varghese v. China Southern Airlines Co. Ltd, 925 F.3d 1339 (11th Cir.2019), does*

[1] (1) Survey of hallucination in natural language generation. Ji Z, Lee N, Frieske R, et al. *ACM Comput Surv.* 2022. (2) Abstracts written by ChatGPT fool scientists. [Preprint] Gao CA, Howard FM, Nikolay S. *bioRxiv.* 2022.

[2] *Mata v. Avianca, Inc.* [Civil Action No: 22 Civ 1461]

indeed exist and can be found on legal research databases such as Westlaw and LexisNexis. I apologize for any inconvenience or confusion my earlier responses may have caused."

"Are the other cases you provided fake."

"No, the other cases I provided are real and can be found in reputable legal databases such as LexisNexis and Westlaw."

12. As can be seen, the lawyer thought that the LLM was engaging in the human process of reading and understanding the question, searching for the correct answer and then communicating the correct answer to the lawyer. In fact, all the LLM was doing was producing outputs (which just happened to be in the form of words) which its mathematical processes related to inputs (which also just happened to be in the form of words). It appears that there has been at least one example in England and Wales where a litigant in person has sought to use ChatGPT in the same way.[3]

13. Of course, it may be unnecessary to add that there are also examples of LLMs being used to manufacture entirely fictitious allegations of misconduct against individuals.[4]

14. **Bias in training data:** Another key risk is inherent in the manner in which an LLM is 'trained'. The fact that the training data is trawled from the internet means that LLMs will inevitably contain biases or perpetuate stereotypes or world views that are found in the training data. There is now a growing body of research on how a range of AI-based tools contain inappropriate biases based on, for example, race and gender. Although the developers of ChatGPT have attempted to put safeguards in place to address these issues, it is not yet clear how effective these safeguards are. Of course, it is also possible to game and manipulate the LLM in certain ways. Ensuring safe and appropriate behaviour from all users can be a significant challenge.

15. **Mistakes and confidential training data:** Finally, ChatGPT and other LLMs use the inputs from users' prompts to continue to develop and refine the system. In consequence, anything that a user types into the system is used to train the software and might find itself repeated verbatim in future results. This is plainly problematic not just if the material typed into the system is incorrect, but also if it is confidential or subject to legal professional privilege.

[3] *Harber v. HMRC* [2023] UKFTT 1007.

[4] We do not identify the examples for obvious good reason, but they are serious and personally damaging.

16. In short, while Generative AI LLM systems have shown impressive capabilities in various natural language processing tasks, they also have come with significant limitations.

Some considerations when using generative AI LLM systems

Practitioners should recognise the constraints and challenges presently embedded in the generative AI LLM software, including:

(1) Possible "hallucinations" and biases (as above-mentioned)

17. The ability of LLMs to generate convincing but false content raises ethical concerns. Do not therefore take such systems' outputs on trust and certainly not at face value. The sanctions on lawyers in the New York case *Mata vs Avianca Airlines Inc* (*supra.*) are a classic example of damage inflicted on a hard-earned reputation because of the court being misled. It matters not that the misleading of the court may have been inadvertent, as it would still be considered incompetent and grossly negligent. Such conduct brings the profession into disrepute (a breach of Core Duty 5), which may well lead to disciplinary proceedings. Barristers may also face professional negligence, defamation and/or data protection claims through careless or inappropriate use of these systems. As set out above, the data used to 'train' generative LLMs may not be up to date; and can sometimes produce responses that are ambiguous, inaccurate or contaminated with inherent biases. Inherent bias may be invisible as it arises not only in the processing or training, but prior to that in the assembling of the training materials. LLMs may also generate responses which are out of context. For these reasons it is important for barristers to verify the output of AI LLM software and maintain proper procedures for checking the generative outputs.

(2) Black Box Syndrome: Lack of explain-ability

18. Like a number of AI tools, generative deep learning AI LLMs are often considered 'heavy black box' models, because it is difficult to understand the internal decision-making processes or provide clear explanations for the output. Some of the software also remains 'proprietary' and therefore confidential. It can sometimes be difficult to interpret the results, due to the multilayer nonlinear model structures and the billions of parameters used. LLMs with Attention Mechanisms[5] may give some ability to see on which parts of the input text the model focuses when generating a response, thereby providing some insights into the decision making. Generative AI LLMs can therefore complement and augment human processes to improve efficiency

[5] Attention Mechanisms allows the model to 'pay attention' to certain elements of the data to give them more weight.

but should not be a substitute for the exercise of professional judgment, quality legal analysis and the expertise which clients, courts and society expect from barristers.

(3) Respect Legal Professional Privilege (LPP), Confidential Information and Data Protection Compliance

19. Be extremely vigilant not to share with a generative LLM system any legally privileged or confidential information (including trade secrets), or any personal data, as the input information provided is likely to be used to generate future outputs and could therefore be publicly shared with other users. Any such sharing of confidential information is likely to be a breach of Core Duty 6 and rule rC15.5 of the Code of Conduct, which could also result in disciplinary proceedings and/or legal liability.

20. Barristers will also need to comply with relevant data protection laws. You should never input any personal data in response to prompts from the system. Note that in March 2023, the Italian Data Protection Authority issued a temporary ban on ChatGPT, largely to investigate whether there was a lack of any legal basis for the collection and processing of any personal data used for training the system, and whether there was a lack of any proper notice to data subjects. Italy, France and Spain are currently investigating OpenAI's processing of data. Using only synthetic data (that is data that is artificially created) on prompts to the LLM represents one possible way to avoid the risk of falling into breach of the General Data Protection Regulation (EU 2016/679) as retained in English law (**UK GDPR**).

21. As practitioners will be aware, the regulatory landscape in this area is in a state of flux and it is difficult to predict exactly what the UK position will be. Under the EU AI Act[6] certain uses of AI tools in legal practice are categorised as 'high-risk' which triggers heightened regulatory obligations. The UK Government's White Paper: A pro-innovation approach to AI regulation[7] published in March 2023, suggests that existing regulators should act in accordance with five principles (similar to the OECD principles on AI[8] although with different wording):

(i) safety, security and robustness;
(ii) appropriate transparency and explain-ability;
(iii) fairness;
(iv) accountability and governance;
(v) contestability and redress.

[6] https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/
[7] http://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1146950/a_pro-innovation_approach_to_AI_regulation_print_ready_version.pdf
[8] https://oecd.ai/en/ai-principles

22. In the UK, the Information Commissioner has published guidance in relation to the development and use of technologies such as ChatGPT: "Generative AI: eight questions that developers and users need to ask"[9].

(4) Intellectual Property (IP) Infringement and Brand Association

23. The precise interaction between the law of intellectual property and LLMs has not yet been the subject of definitive consideration. However, barristers will need to critically assess whether content generated by LLMs might violate intellectual property rights, especially third-party copyright. As a sizable amount of text data, such as books, papers, and other written materials were used to train ChatGPT and other LLMs, it is clearly possible that content produced may violate copyright or other IP rights in previously published materials. Several IP claims against generative AI owners have been lodged for allegedly unlawful copying and processing of millions of copyright-protected images, and associated metadata.[10]

24. Further, one should be careful not to use, in response to system prompts, words which may breach trademarks or give rise to a passing-off claim. Often the terms of service of a LLM give the company owning the LLM tools unlimited use of information given to the system.

Professional considerations

25. Irresponsible use of LLMs can lead to harsh and embarrassing consequences, including claims for professional negligence, breach of contract, breach of confidence, defamation, data protection infringements, infringement of IP rights (including passing off claims), and damage to reputation; as well as breaches of professional rules and duties, leading to disciplinary action and sanctions.

26. There is a growing body of material in which practitioners and others discuss their use of LLMs in the course of legal practice. This guidance is concerned only to explain some of the pitfalls. It is for barristers themselves to work out how and in what context a LLM might assist them in providing legal services. This process is likely to be a changing one as the technology itself develops as it is doing and with increasing speed.

---

[9] https://ico.org.uk/about-the-ico/media-centre/blog-generative-ai-eight-questions-that-developers-and-users-need-to-ask/

[10] Cases such as (1) Getty Images against Stability AI Inc. for copyright infringement in AI training data; (2) Class actions in US against OpenAI challenging ChatGPT by Paul Tremblay and Mona Awad, and Sarah Silverman, Christopher Golden and Richard Kadrey and others suing OpenAI and Meta.

27. Barristers should also keep abreast of relevant Civil Procedure Rules, which in the future may implement rules/practice directions on the use of LLMs; for example, requiring parties to disclose to the court when they have used generative AI in the preparation of materials. This approach has already been adopted by the Court of the King's Bench in Manitoba.[11]

Conclusion

28. In conclusion, technical progress and the pressures of competition may lead to the increasing adoption of AI, including LLMs. The best-placed barristers will be those that make the effort to understand these systems and, if appropriate, use them as tools in their practice, while maintaining control and integrity in their use. There is nothing inherently improper about using reliable AI tools for augmenting legal services; but they must be properly understood by the individual practitioner and used responsibly, ensuring accuracy and compliance with applicable laws, rules and professional codes of conduct.

Important Notice

This document and sample policy has been prepared by the Bar Council to assist barristers and chambers on matters of information security. **It is not "guidance" for the purposes of the BSB Handbook I6.4, and neither the BSB nor bodies regulating information security, nor the Legal Ombudsman is bound by any views or advice expressed in it.** It does not comprise - and cannot be relied on as giving - legal advice. It has been prepared in good faith, but neither the Bar Council nor any of the individuals responsible for or involved in its preparation accept any responsibility or liability for anything done in reliance on it. For fuller information as to the status and effect of this document, please see here.

[11] https://www.lawgazette.co.uk/news/canadian-judges-demand-to-know-if-ai-used-in-submissions/5116452.article