



Oifig an Príomh Bhreitheamh
Office of the Chief Justice

GUIDELINES FOR JUDGES ON THE RESPONSIBLE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

June 2024

AI tools are now common in daily life and have been used for some time without difficulty by the judiciary and parties appearing in Irish courts, for example, search engines used to conduct research in specialist legal databases or machine learning technology used to identify potentially relevant documents in large discovery projects.

These guidelines relate to a particular form of AI known as generative artificial intelligence (GenAI). GenAI represents a new area of innovation because this technology creates new content, including text, images, audio, and computer code. GenAI allows the user not merely to find, sort or index materials that already exist, but to create entirely new content in response to questions or prompts. For that reason, GenAI presents new challenges and opportunities for all users, including the judiciary. Publicly available GenAI chatbots include ChatGPT, Bing Chat or Google Bard. Many more platforms are in development.

It is essential for all judges to understand the benefits and limitations of GenAI which are set out in these guidelines. Judges must ensure that any use of GenAI tools is consistent with the overriding obligation to ensure the integrity and reliability of the legal research and analysis which underpins judicial decision making. Any use of GenAI tools by the judiciary must be consistent with the judiciary's overarching obligation to protect the independence and integrity of the administration of justice and the protection of fundamental rights. Judges remain responsible for their decisions

These guidelines provide assistance to judges on how to identify and manage the use of GenAI by others (including lawyers and unrepresented parties) in court proceedings. Examples of potential uses of GenAI are set out – in summary it may be useful to use public GenAI tools to assist you with aspects of your work such as speeches or administrative or routine tasks, but it is not recommended to use such tools for core judicial work such as conducting research to find new information which you cannot verify independently or for legal analysis. The use of GenAI on trusted databases can be a useful supplementary tool in some contexts but it cannot replace the judgment, empathy, and ethical responsibility of judges in the performance of their judicial role.

These guidelines adopt a similar approach to judicial guidelines in comparator jurisdictions. The guidelines will be updated from time to time to reflect relevant technological advancements and developments. An explanation of common terms used in these guidelines as well as a set of FAQs are included.

GUIDELINES FOR JUDGES ON THE RESPONSIBLE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

1) GenAI: How It Works & How It Can Go Wrong

Before using GenAI chatbots ensure you have a basic understanding of their capabilities and limitations.

Some key limitations:

- The output generated by a GenAI chatbot is what it predicts to be the most likely combination of words (based on the documents and data that it holds as source information). So, even if the output looks convincing, it may not be factually correct.
- Public GenAI chatbots (like ChatGPT) cannot be restricted to provide answers solely from authoritative databases, or databases with any significant training data on Irish law. Rather, such GenAI chatbots' responses to prompts typically draw on a multitude of internet sources, many of which are uncured and potentially unreliable, to generate new text as a potential answer to the user's query. The technology generates the response using a complex algorithm based on the prompts received and the data with which they have been "trained" – namely the vast amount of content on the open web with its varying degrees of accuracy.
- As with information generally available on the internet, GenAI tools may be a useful shortcut to find material you would recognise as correct but do not have to hand. However, when seeking new information, material generated by GenAI tools must be verified independently in terms of its accuracy and comprehensiveness before you can rely on it. The quality of any answers you receive will depend on how the GenAI chatbot has been trained, the reliability of the training data, and how you engage with the relevant GenAI chatbot, including the "quality" of the prompts you enter. Its answers may be skewed by context – for example, by information on the law in other jurisdictions which may not apply in Ireland. GenAI tools do not have a human perspective from which to provide judgment and nuance. They cannot understand material although they may appear to. The responses they provide will be generated without any understanding or analysis and (depending on the specific prompts given to it) the answers provided may not be tailored to account for the unique facts of a specific case. They do not have a bias detector and have no embedded fact-checking capabilities.
- Even with the best prompts, output which superficially appears convincing may be seriously inaccurate, incomplete, misleading, or biased.
- Public AI chatbots are not confidential. The information you enter could potentially be retained for purposes such as further training material. Any information you input into a public AI chatbot, even for preliminary prompts or questions should be treated as though published to all the world.

- You must not enter any private, confidential, or legally privileged information into a public AI chatbot (even one you hold a subscription for).

2) GenAI is Neither Confidential Nor Private

Unless otherwise guaranteed, assume that any material entered into a GenAI system is going into the public domain. Do not enter any information that is private, confidential, sensitive, suppressed, or legally privileged or which is personal data or data that could potentially identify any party.

- Some GenAI chatbots can remember every question asked, as well as any other information you input. That information could then be used or repeated in response to queries from other users. As a result, your enquiries of a GenAI chatbot could become publicly known. This could breach data protection and privacy rights, court orders or statutory prohibitions on publication. If private or confidential material enters the public domain through the use of GenAI, a judge will not have control over how that information is used.
- In the event of unintentional disclosure of private, confidential, suppressed, or privileged information:
 - Judges should notify the List Head and List Heads should inform their President.
 - All judges should immediately report breaches to both the supervisory authority for courts acting in their judicial function under the [section 157 of the Data Protection Act 2018](#), and also to the Courts Service Data Protection Officer.
- Learn how to limit the information you input into sources used by AI. You should disable the chat history in AI chatbots if this option is available. Some AI platforms, particularly if used as an App on a smartphone, may request various permissions which give them access to information on your device. In those circumstances you should refuse all such permissions.
- The use of GenAI may raise copyright issues. Compliance with the law of copyright is the responsibility of the judge.

3) Ensure Accountability & Accuracy

You must independently check and verify the accuracy of any information provided to you by a GenAI chatbot before you rely on it. This includes a responsibility to check references and citations to primary sources.

- Information provided by GenAI chatbots may be inaccurate, incomplete, or out of date. It may be based on foreign law that does not apply in Ireland, although it purports to represent Irish law. It may not be possible to secure transparency as to how

the GenAI chatbot has provided its response. This is why independent verification of information provided is essential.

- GenAI chatbots may:
 - Cite fictitious cases, make up citations or quotations, or refer to legislation, articles or legal texts that do not exist;
 - provide incorrect or misleading information on Irish law or how it might apply;
 - make factual errors; and
 - confirm that information is accurate if asked, even when it is not.

4) Be Aware of Ethical Issues

Have regard to ethical issues – particularly biases and the need to address them.

- Chatbots generate responses based on the dataset on which they are trained (generally this is information from the internet). Information generated by a GenAI chatbot will reflect all the biases or misinformation in its training data. You should always have regard to this possibility and the need to correct it.
- Early iterations of GenAI displayed gender and racial biases, sometimes to an extreme degree. It is hard to overstate the danger of unchecked GenAI material being used by the judiciary.
- These ethical concerns are closely related to the confidentiality and privacy issues already highlighted.

5) Maintain Security

Follow best practices for maintaining your own and the court's security.¹

For example:

- Use work devices (not personal devices) to gain access to AI chatbots for work purposes. This is because your work devices have enhanced and consistent security protections in place and are fully backed up should there be a need to retrieve information.
- Use your work email address.
- Ensure you have selected a legitimate site or an official mobile device app. There are many unofficial sites and apps available that could be malicious.
- switch off chat history or regularly delete interactions to limit potential data breach exposure.

- If you have a paid subscription to a GenAI chatbot platform, use it. Paid subscriptions are generally more-secure – but the inherent risks and limitations of GenAI chatbots remain.
- A useful reference published by our [National Cyber Security Centre](#) is informative.
- If there has been a potential security breach while you are using a Courts Service issued device, please immediately report this potential security breach to the Courts Service IT Helpdesk.

6) **All Use Of GenAI Must Be Disclosed To You By Judicial Support Staff Assisting You On Any Matter**

Judicial support staff, including all Judicial Assistants, must disclose to their supervising judge if they have used GenAI chatbots (or any other GenAI tools) and the steps they have taken to mitigate any risks.

- Separate guidelines in similar terms to these guidelines will be issued by the Courts Service to all judicial support staff. Those guidelines will highlight the need to disclose the use of GenAI by judicial support staff. This is to ensure that judges are alerted to the need for particular vigilance and scrutiny of any information produced in this manner. Judges are personally responsible for all material on which they rely.

7) **Be Aware Of GenAI Use By Others**

Check the accuracy of information contained in submissions or other documents that show signs they were produced by a GenAI chatbot.

- All legal representatives are responsible for the accuracy of the material they put before the court. Barristers and solicitors have a professional obligation as officers of the court and this extends to taking reasonable steps to satisfy themselves of the accuracy and reliability of material, including citations, which they put before the court. Provided AI is used responsibly, and the output is appropriately tested and verified there is no reason why a legal representative ought to refer to its use, but this depends on context. Judges may request confirmation that legal professionals have independently verified the accuracy of any research or case citations generated by a GenAI chatbot if a judge has any concerns regarding the materials provided to the court.
- Unrepresented parties often use GenAI chatbots to identify and explain relevant laws and legal principles or to prepare basic legal documents in proceedings. Indeed, GenAI chatbots may be the only source of advice or assistance some unrepresented parties receive. Unrepresented parties or McKenzie friends will often not have the skills or training to independently verify the legal information provided to them by a GenAI chatbot. They may not appreciate the significant limitations and risks involved in using such information or that GenAI chatbots are not a substitute for a legal

representative and cannot give tailored advice. If it appears that a GenAI chatbot may have been used to prepare submissions or other documents by an unrepresented party, a judge may need to inquire whether a GenAI chatbot has been used and in what manner, and to ask what checks for accuracy have been undertaken (if any).

- “Red flags” that may indicate a GenAI chatbot has been used include:
 - submissions that use American spelling or refer to foreign cases;
 - content that (superficially at least) appears to be highly persuasive and well-written, but on closer inspection contains obvious substantive errors;
 - references to cases that do not sound familiar;
 - parties citing different bodies of case law in relation to the same legal issues;
 - confused concepts – as where a particular term (such as disclosure) has alternative meanings in different contexts and the submissions are not focussed on the relevant issue; and
 - submissions that do not accord with your general understanding of the law in the area.
- Be aware that GenAI can also fabricate convincing images, audio, video and other media, which parties could present as evidence. Judges should be aware of this possibility and the potential challenges posed by deepfake technology. It is appropriate for judges to question the use of GenAI as part of their consideration of the validity of any evidence produced by parties.

EXAMPLES OF POTENTIAL USES OF GenAI AND USES NOT RECOMMENDED

POTENTIAL USES:

Summarising information:

GenAI chatbots can be useful in condensing or summarising information. However, the following should be considered:

- Using a GenAI chatbot to summarise a portion of a textbook or other intellectual property could breach the author's copyright.
- Any such use would need to be carefully reviewed to ensure the summarised passage carries the same meaning as the original content.
- Depending on context, the source may need to be acknowledged and citations added.

Speech writing:

GenAI chatbots can be a helpful tool in planning a speech and producing an outline of potential speaking points. It could then be used to elaborate further on potential content for a specific speaking point. However:

- Careful oversight by a judge is required to use GenAI chatbots responsibly in this way.
- It is important to ensure that any AI-generated material is accurate and supported by reliable sources.

Administrative tasks:

Using GenAI chatbots for administrative tasks such as the drafting of emails requires caution:

- Remember that GenAI chatbots can retain any data put into them, including names, email addresses and so on. They could potentially disclose such information to a subsequent user.
- GenAI chatbots lack nuance and understanding of context. As a result, any AI-generated written material needs to be reviewed for appropriate use of tone, language, cultural appropriateness, etc.

TASKS NOT RECOMMENDED:

Legal research:

With straightforward areas of law or for material you would recognise as correct but do not have to hand, GenAI chatbots might be able to produce helpful, higher-level legal explanations or summaries of relevant legal principles. However, they should not be used as a means of conducting research to find new information which you cannot verify independently.

- Care must be taken to provide accurate and reliable information to the GenAI chatbot.
- The quality of legal research produced will be influenced by the type of prompts the GenAI chatbot is given — it is only capable of producing material in response to what it is asked. Writing high quality prompts, and tailoring them in response to the answers received, is a skill that takes some practice.

- While GenAI chatbots can produce some high-quality outputs (for example identifying and explaining relevant legal principles) they are also prone to errors and “hallucination”. For example, it is currently common for GenAI chatbots to generate fictitious legal citations. In doing so, the GenAI chatbot may appear plausible and authoritative but the output may be seriously misleading.
- The use of GenAI chatbots is not a substitute for conducting research using trusted sources such as academic texts or legal databases.
- GenAI chatbots are limited by the date range, jurisdictional information, and type of legal materials they can access. For example, the outputs generated are often based on publicly available US legal materials, reflecting the data set that GenAI chatbots are trained on (generally, the internet).

Legal analysis:

Using GenAI chatbots for legal analysis is not recommended because:

- GenAI is ill-suited to legal analysis as it generates text based on probability, rather than an understanding of text’s content or human inferences.
- GenAI does not have an ability to critically examine the patterns it identifies in data — this can result in it drawing inaccurate or biased conclusions.
- Current public GenAI chatbots do not currently produce convincing analysis or reasoning. To rely on it uncritically would compromise the integrity of the judicial process.

COMMON TERMS

- **Artificial Intelligence (AI):** the theory and development of computer systems able to perform tasks normally requiring human intelligence.
- **Bing Chat:** Microsoft's GenAI chatbot
- **ChatGPT:** OpenAI's GenAI chatbot — "Chat **G**enerative **P**re-**T**rained **T**ransformer"
- **Generative AI (GenAI):** a form of AI that enables users to quickly generate new content — can include text, images, sounds and other data. Some GenAI tools are designed to take actions.
- **GenAI chatbot:** computer programme for simulating online human conversations using generative AI.
- **Google Bard:** Google's GenAI chatbot.
- **Large Language Model (LLM):** an AI algorithm which learns to predict the next best word or part of a word in a sentence. Generative AI chatbots generally use LLMs to generate responses to "prompts".
- **Prompt:** instructions or questions provided by a user to a generative AI chatbot to obtain the desired answer/output.

FAQs

What are GenAI Chatbots and how do they work?

GenAI chatbots, such as ChatGPT, Google Bard and Bing Chat, are AI tools that have been trained to respond in a conversational, online chat style. Users can enter prompts (questions) to get the GenAI chatbot to do things like generate or summarise text or answer questions. They can then enter more prompts to refine the chatbot's response.

GenAI chatbots are built on AI algorithms called Large Language Models (LLMs). LLMs analyse a large amount of training text to predict the probability of the next best word in a sentence given the context. Just as Google offers to autocomplete your search, LLMs autocomplete repeatedly to form words, sentences, and paragraphs of text.

LLMs have been further trained on ideal human written responses to prompts, and on survey results, about which responses sound most natural or best mimic human dialogue.

What information is used to generate the response I receive?

As GenAI chatbots output the next best word based on the text contained in their LLMs dataset, the source, quality, diversity, and any bias of that text will determine the responses received.

Generally, the text used to train GenAI comes from various internet sources, such as webpages, online books, and social media posts. This impacts the accuracy of any legal information a GenAI chatbot may provide, although legal information on the internet, such as legislation and some legal commentary, may form part of the dataset.

GenAI chatbots cannot distinguish between facts, inferences and opinions contained in its dataset. This means that text generated by GenAI chatbots may contain incorrect, opinionated, misleading, or biased statements presented as fact. The training data used by some chatbots may also be out of date.

GenAI draws on data that has been previously inputted by users. That means it may, depending on the platform and privacy settings, retain anything that users enter as prompts.

In terms of bias and harmful or dangerous information, algorithmic countermeasures have been employed to reduce the chance that such data ends up being used in GenAI chatbot responses. However, these countermeasures do not eliminate the risk of biases, or harmful or dangerous information being included in responses.

Why am I receiving incorrect information?

As GenAI chatbots base their responses on probability-derived calculations about the next best word in context, they are unable to reliably answer questions that require a nuanced understanding of language content. These tools have no intrinsic understanding of what any word they output means, nor a conception of truth.

GenAI chatbots are known to produce false information that may appear true — this is called “hallucinating”. In the legal field, this may lead to GenAI chatbots making up non-existent cases, changing facts, and producing incorrect citations. By their very nature, LLMs will produce answers that sound convincing even when they are entirely fictitious.

Multimedia-generative AI

Multimedia-generating AI, such as DALL-E, Midjourney and Stable Diffusion, have emerged alongside text generative AI. These tools have been trained so that the algorithm can best predict the patterns and styles associated with different multimedia forms and their text descriptions. They are then able to generate new multimedia (i.e., images, audio, or video) that is similar in style and content.

This technology allows for the creation of multimedia that appear authentic but are fake. These multimedia are known as ‘deepfakes’ and can be defamatory or spread dangerous misinformation or disinformation. Concerns may arise over the authenticity of multimedia when used in evidence.