SUPREME COURT
OF QUEENSLAND

Australian Insurance Law Association (AILA) National Conference
Thursday, 12 September 2024
Gold Coast Convention Centre

NAVIGATING THE AI ABYSS: ETHICAL CHALLENGES IN THE AGE OF GENAI

Chief Justice Helen Bowskill[1]

Ladies and gentlemen, esteemed insurance lawyers, and fellow legal aficionados.

As we stand at the precipice of a technological revolution, our legal landscape is undergoing a seismic shift. The emergence of **Generative AI** – a powerful force that breathes life into code, conjuring prose, images, and even legal arguments – has thrust us into uncharted territory. But with great power comes great responsibility, and today, we delve into the **ethical conundrums** that accompany this digital sorcery.

Let's try that again, as me this time, rather than Microsoft Copilot, to whom (which?) I give credit for the catchy title and intro.

Members of the Australian Insurance Law Association, lawyers and others working within the insurance industry, I acknowledge all of you. I also acknowledge the first owners and custodians of this land, and the land across Queensland. I commend their ancestors and elders, for their patience, courage and wisdom.

You have come to the right place for a conference themed Sunny Side Up. What a wonderful location to hold this annual conference, which presents an important opportunity for exchanging ideas, networking and just simply taking some time out to allow yourself to think about the bigger picture in a way that the grind of our day to day working lives sometimes does not allow.

---

[1] With thanks to Ms Alicia George, my research assistant, and to my associate, Ms Bronte Donohoe for their assistance in preparing this presentation; and, to a lesser extent, Microsoft CoPilot.

It is in that context that I am to say something about the challenges that arise from the use of the generative AI in the context of court proceedings. I am hoping to leave you seeing the sunny side of AI; rather than thinking of it as an abyss. In that regard, I endorse the observation made by Sir Geoffrey Vos, Master of the Rolls and Head of Civil Justice in the UK, in a speech he gave in March 2024 to the Manchester Law Society, that:

> "… there is nothing scary about AI. It is just a technological tool that has, by the way, been around for years. You use it happily every time you pick up your smart phone.
>
> What is scary, as always, is a very small number of ill-intentioned people. Such people might use AI inappropriately if we do not protect ourselves properly, and build in human controls. But that is not really any different to other technological developments that history has produced. Cars, aeroplanes, industrial machinery, oil, mining and almost every other technological innovation can be very dangerous to people, and even to humanity itself, if misused."

So, what are we talking about when we refer to AI? Well, we are really talking about generative AI, as opposed to simply AI, because generative refers to the form of AI which enables users to generate new content (images, text etc). A generative AI chatbot is simply a computer program which simulates online human conversations using generative AI.

I provide the following brief outline to put my observations into context. It is taken from guidelines recently published by Queensland courts and tribunals for non-lawyers.

Despite the name, generative AI chatbots are not actually intelligent in the ordinary human sense. Nor is the way in which they provide answers analogous to the human reasoning process. It is important to note:

- generative AI chatbots are built on Large Language Models, or 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.

- This means the answers which generative AI chatbots generate are what the chatbot predicts to be the most likely combination of words (based on the documents and data that it holds as source information), not necessarily the most accurate answer.

- And because their responses are based on probability-derived calculations about the next best word in context, these tools are unable to reliably answer questions that require a nuanced understanding of

language content. They have no intrinsic understanding of what any word they output means, nor a conception of truth.

The answers provided by generative AI chatbots depend on the content of the datasets from which they are trained. You should note the following limitations:

- Generally, the text used to train public generative AI chatbots comes from various internet sources, such as webpages, online books, and social media posts. It does not necessarily come from authoritative or up to date databases.

- The current public generative AI chatbots appear to have had limited access to training data on Australian law or the procedural requirements that apply in Australian courts and tribunals. Even when that improves, there will be a limitation based on the currency of the data on which they have been trained.

- Generative AI chatbots cannot distinguish between facts, inferences and opinions contained in their source datasets. This means that text which they generate in response to a prompt may contain incorrect, opinionated, misleading or biased statements presented as fact.

And the quality or helpfulness of any answers you receive from generative AI chatbots depends on the questions or "prompts" which you ask.

Returning to Copilot – when I asked it to help me with an introduction for a speech to lawyers about the ethical challenges associated with generative AI, it included an identification of what it called the "murky corners". There are five of them:

1. The Hallucinating Oracle: Output Risks
2. Bias: The Ghost in the Machine
3. Copyright Quandaries: The Art of Plagiarism
4. The Fine Line: Human-AI Collaboration
5. The Regulatory Maze: Navigating Uncertainty

Let me say something about each of these – but mostly about the first one.

1. The Hallucinating Oracle: Output Risks

We have all heard about this. Information provided by generative AI chatbots may be inaccurate, incomplete, or out of date. It may also be based on overseas law that does not apply in Australia – although this should change with the introduction of the LexisNexis and ThomsonReuters AI products. But putting such products to one side, generative AI chatbots can:

- make up fake cases, citations and quotes, or refer to legislation, articles or legal texts that do not exist;

- provide incorrect or misleading information about the law or how it might apply in a case; and

- get facts wrong.

Fake material produced by generative AI chatbots can seem like it has been taken from a real source even when it has not.

One case example in which this has arisen is *Mata v Avianca Inc* 678 F Supp 3d 443 (a 2023 decision of the United States Southern District Court, New York).[2]

A passenger on an aeroplane brought a claim against the air carrier for damages for personal injuries after the metal serving cart struck his knee. As I read the decision, the air carrier filed a strike out application. The lawyers for the passenger filed submissions in opposition to that. There was a problem with the submissions. The air carrier's lawyers said they could not find the cases that had been cited in it. The Court could not find them either – and ordered the passenger's lawyers to file an affidavit attaching copies of them. The problem was, the lawyers had used ChatGPT to prepare the submissions, and it included fake cases, with fake quotes and citations. As if that was not bad enough, the lawyers doubled down and stood by the fake cases after being ordered to provide copies.

Judge Castel commenced his reasons by observing:

> "In researching and drafting court submissions, good lawyers appropriately obtain assistance from junior lawyers, law students, contract lawyers, legal encyclopedias and databases such as Westlaw and LexisNexis. Technological advances are commonplace and there is nothing inherently improper about using a reliable artificial intelligence tool for assistance. But existing rules impose a gatekeeping role on attorneys to ensure the accuracy of their filings. [The passenger's lawyers in this case] abandoned their responsibilities when they submitted non-existent judicial opinions with fake quotes and citations created by the artificial intelligence tool ChatGPT, then continued to stand by the fake opinions after judicial orders called their existence into question."

It was noted that, if the lawyers had "come clean" once the issue was first pointed out, the outcome would have been very different. But as they did not, a finding was made that they had acted in bad faith, "based upon acts of conscious avoidance and false and misleading statements to the court". The sanction imposed was a financial penalty; as well as a requirement to write to not only the passenger himself, ie the client, advising him of the decision; but also to write to all the judges who had been improperly identified in the fake cases cited.

Judge Castel went on to say:

> "Many harms flow from the submission of fake [case authorities]. The opposing party wastes time and money in exposing the

[2] *Mata v Avianca Inc*, 678 F Supp 3d 443 (SD NY, 2023).

> deception. The Court's time is taken from other important endeavours. The client may be deprived of arguments based on authentic judicial precedents. There is potential harm to the reputation of judges and courts whose names are falsely invoked as authors of the bogus opinions and to the reputation of a party attributed with fictional conduct. It promotes cynicism about the legal profession and the American judicial system. And a future litigant may be tempted to defy a judicial ruling by disingenuously claiming doubt about its authenticity."

Now we'd all like to think that is an extreme and outrageous case, which could never happen here.[3] Let us hope so. But it is a stark example and reminder of the need to check the output of anything produced with generative AI to ensure accuracy.

A similar issue arose in a case in the Supreme Court of British Columbia, *Zhang v Chen* [2024] BCSC 285, which involved a lawyer citing non-existent cases in a document filed in court in relation to an application for parenting orders.[4] That lawyer did "come clean" and apologise to the opposing party's lawyer, as soon as the error was identified, which was before the matter was heard in court. An application was made for the lawyer to personally pay the successful party's costs of the application, and also to pay "special costs" (which I infer is indemnity costs). Concerningly, the judgment records that the lawyer was not aware of the risks of ChatGPT producing fake cases – despite having a PhD and the publication of notices by the Law Society, let alone the warnings on the ChatGPT website itself.

Whilst Masuhara J described what occurred as "alarming", he was satisfied the lawyer had no intent to deceive. His Honour observed, at [29], that:

> "Citing fake cases in court filings and other materials handed up to the court is an abuse of process and is tantamount to making a false statement to the court. Unchecked, it can lead to a miscarriage of justice."

However, his Honour noted that the error was identified before the court hearing and, having regard to the substantial legal teams on both sides, "there was no chance here that the two fake cases would have slipped through". Being satisfied there was no intention to deceive, his Honour did not consider the circumstances justified the imposition of a "special costs" order against the lawyer. However, since the insertion of fake cases resulted in delay and "additional effort and expense" that was wasted, the lawyer was ordered to personally pay, effectively, the costs "thrown away" as a consequence. She was also ordered to review all of her files before the Court and check case citations and summaries in any documents filed.

---

3 Postscript: Following the presentation of this speech, I became aware of an Australian case in which non-existent cases were cited to a court: *Handa & Mallick* [2024] FedCFamC2F 957, and the related decision, resulting in an order for referral of the solicitor to the Victorian Legal Services Board, in *Dayal* [2024] FedCFamC2F 1166.

4 *Zhang v Chen* [2024] BCSC 285.

Interestingly, Masuhara J says the following, at [38]:

> "The risks of using ChatGPT and other similar tools for legal purposes was recently quantified in a January 2024 study: Matthew Dahl et. al., "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models" (2024) …. The study found that legal hallucinations are alarmingly prevalent, occurring between 69% of the time with ChatGPT 3.5 and 88% with Llama 2. It further found that large language models ('LLMs') often fail to correct a user's incorrect legal assumptions in a contrafactual question setup, and that LLMs cannot always predict, or do not always know, when they are producing legal hallucinations. The study states that '[t]aken together, these findings caution against the rapid and unsupervised integration of popular LLMs into legal tasks.'"

By way of final comment, his Honour said this, at [46]:

> "As this case has unfortunately made clear, generative AI is still no substitute for the professional expertise that the justice system requires of lawyers. Competence in the selection and use of any technology tools, including those powered by AI, is critical. The integrity of the justice system requires no less."

Closer to home, a sentencing decision given earlier this year by Mossop J of the ACT Supreme Court highlights a problem when generative AI is used to produce evidentiary material – in that case, a personal character reference for the offender. In *DPP v Khan* [2024] ACTSC 19, Mossop J referred to personal references that were tendered in support of the offender.[5] His Honour formed the view that one of the references, which was said to be from the offender's brother, was written with the assistance of something like ChatGPT. There were two things that stood out. First, the way in which the author's relationship with the offender was introduced, which was not what you would expect from a brother:

> "I have known Majad both personally and professionally for an extended period, and I am well-acquainted with his unwavering commitment to his faith and community."

The second was a paragraph, praising the offender's commitment to cleanliness, which contained "non-specific repetitive praise", as follows:

> "Majad's commitment to cleanliness and order is another facet of his character that stands out. He maintains a meticulous approach to his surroundings, expressing a strong aversion to disorder. His proactive attitude towards cleaning, both inside the house and in the community, reflects a sense of responsibility and respect for the environment. His efforts extend to keeping the streets and driveways clean, a testament to his commitment to a well-maintained and orderly community."

---

5 *DPP v Khan* [2024] ACTSC 19.

In the end, Mossop J found it difficult to assess the weight to be given to the reference, and commented that it is "clearly inappropriate that personal references used in sentencing proceedings are generated by" AI. Mossop J considered there was a positive duty on counsel appearing at sentencing proceedings to make appropriate enquiries prior to tendering and be able to inform the court as to whether any character references had been prepared using generative AI.

By extrapolation – what about affidavits or witness statements? The same point applies, but with even more serious potential consequences than matters of weight, where a person swears on oath or by affirmation as to the accuracy of the statement.

What about other evidence? Or tests or experiments or recreations – is it ok to use generative AI? Some of the possible considerations are outlined in a decision of the Superior Court of Washington in the matter of *State of Washington v Joshua Puloka (aka Joshua EverybodyTalksAbout)* (from March 2024).[6] The defendant was accused of shooting five people, killing three, outside a Seattle bar. A civilian witness recorded a video of some part of the incident on their iPhone. That original recording was about 10 seconds. The defence sought to tender an AI enhanced version of the video prepared by a "self-identified videographer and filmmaker" who was not a forensic video technician.

The AI version was described as having enhanced the original, including by "adding clarity". The evidence included that the AI tool used "'machine learning', employing specific processing models based on a vast library of videos", but the videographer did not know what videos the AI-enhancement models were trained on and did not know whether the models employed generative AI. The prosecution's forensic video analyst gave evidence that the AI enhanced video increased the number of pixels in the video, and "created false image detail" – meaning, it had the effect of changing the meaning of portions of the video.

The use of AI tools to enhance video introduced in a criminal trial was described as a "novel" technique, which had not achieved general acceptance in the relevant scientific community. The AI enhanced video was held not to be admissible, due to its potential to misrepresent the events. The original video, blurry as it was, remained the best evidence.

Anyone using generative AI must take care to ensure the content of what is produced is accurate. This includes self-represented litigants.

An example of the use of generative AI in a court proceeding is *Yousseff v Eckersley* [2024] QSC 35.[7] In that case, the self-represented plaintiff prepared his written submissions, following a trial, using ChatGPT. He told the judge he had done that, and "vouched for the accuracy of his submissions", although

---

6 *State of Washington v Joshua Puloka (aka Joshua EverybodyTalksAbout)* (from March 2024).

7 *Yousseff v Eckersley* [2024] QSC 35.

"stated that this platform assisted in their organisational structure and added a flourish to his submissions" (at [17]).

We expressly tell non-lawyers in our guidelines, that they must check the accuracy of any information they get from a generative AI chatbot before using that information in court or tribunal proceedings.

A self-represented litigant in a case in the Northern Ireland High Court disclosed their use of ChatGPT. They had apparently used it to answer a series of questions, criticising counsel, solicitors and judges, and then used those answers in support of their case, suggesting the answers had added weight *because* they were produced by artificial intelligence, which "does not have personal opinions, beliefs or feelings". As the judge observed, though, "[s]adly Chat GPT seemed unable to recognise or correct the misuse by Mr Carlin in one of his questions of the phrase 'cast dispursions' rather than "cast aspersions". See *Santander UK Plc v Carlin* [2023] NICh 5.[8]

In another example from the UK, *Harber v The Commissioners for His Majesty's Revenue and Customs* [2023] UKFTT 01007 (TC), the self-represented appellant had been penalised for failing to notify her liability to capital gains tax.[9] She sought to appeal the penalty on the basis that she had a reasonable excuse, because of her mental health condition and/or because it was reasonable for her to be ignorant of the law. She filed a submission, citing a number of decisions in which an appellant had been successful in showing a reasonable excuse existed. But none of the decisions existed – they were all generated by artificial intelligence. The Tribunal accepted that the self-represented appellant was unaware the AI cases were not real and did not know how to check their validity. The Tribunal just put the submission to one side.

What about judges?

An American judge, from the 11th Circuit Court in the United States Court of Appeals, used ChatGPT and Google Bard (now Gemini) to ascertain the ordinary meaning of the word "landscaping". In *Snell v United Specialty Insurance Company* (11th Cir, No 22-12581, 2024), the plaintiff was a landscaper the subject of a claim arising from alleged negligent installation of a ground-level trampoline in a client's backyard.[10] He sought indemnity from his insurer. The insurer denied indemnity, saying the claim fell outside the scope of the policy. The question was whether installation of such a trampoline fell within the common understanding of the term "landscaping" as used in the insurance policy. The Court held that it did not. Rather than that conclusion turning on the meaning of "landscaping", it was actually driven by a "quirk" of Alabama law – that a person's insurance application forms part of the policy – and the fact that Mr Snell had expressly denied in his application that his work included any recreational or playground equipment construction or erection.

---

8 *Santander UK Plc v Carlin* [2023] NICh 5.

9 *Harber v The Commissioners for His Majesty's Revenue and Customs* [2023] UKFTT 01007 (TC).

10 *Snell v United Specialty Insurance Company* (11th Cir, No 22-12581, 2024).

Those two things taken together meant the policy did not cover the work in any event.

One of the members of the Court, Newsom J, wrote a short concurring judgment, as he said, "simply to pull back the curtain on the process by which I thought through one of the issues in this case" and "to make a modest proposal regarding courts' interpretations of the words and phrases used in legal instruments". As he said, the "off-ramp" – ie the actual legal answer – wasn't always obvious to him, and he spent "hours and hours (and hours)" labouring over the question of the ordinary meaning of "landscaping". And that got him thinking that maybe ChatGPT could assist. In a delightful narrative, the judge tells the story of how he used ChatGPT, and checked his answers on another platform, Google Bard (now Gemini). That lead him to pose the question, "might LLMs be useful in the interpretation of legal texts?", as one implement among several in the textualist toolkit. He outlined a number of "pros" and "cons" to doing that. In the interests of time, I won't go through all of those, it is quite a lengthy discussion, but I'll read you his concluding paragraph:

> "In his most recent year-end report on the state of the federal judiciary, Chief Justice Roberts cautioned that the 'use of AI requires caution and humility.' Roberts, *supra*, at 5. I wholeheartedly agree. Importantly, though, I also agree with what I take to be the report's assumption that AI is here to stay. Now, it seems to me, is the time to figure out how to use it profitably and responsibly. It's in that spirit that I've offered *these* preliminary thoughts about whether and how LLMs might aid lawyers and judges in the interpretive enterprise. Plenty of questions remain, and I'm sure I haven't even identified all of them. But – and this is my bottom line – I think that LLMs have promise. At the very least, it no longer strikes me as ridiculous to think that an LLM like ChatGPT might have something useful to say about the common, everyday meaning of the words and phrases used in legal texts.
>
> Just my two *cents*."

Another judge who disclosed his use of ChatGPT when preparing a judgment is Lord Justice Birss of the UK Court of Appeal. In a speech he gave in September 2023, Birss LJ said he had used ChatGPT to provide a summary of an area of law. He said he received a paragraph that he considered was "jolly useful". Importantly, it was an area of law his Honour knew well, therefore he was well placed to assess the accuracy of the summary. And as he said, he took full personal responsibility for the content, and simply used the chatbot to do a summarising task that he would otherwise have done himself.

Further afield, there are examples of judges using ChatGPT to answer more substantive legal questions. For example, a judge in Columbia apparently disclosed in a judgment his use of ChatGPT as part of the process of determining whether an autistic child's insurance should cover all the costs of his medical treatment. He posed substantive questions like "is an autistic child exempt from co-payments for therapy?" and "has the jurisprudence of the

constitutional court made favourable decisions in similar cases?" The judge did emphasise that any information provided by the AI tool was fully fact checked and the purpose was to assist and speed up the process, not to replace judicial expertise.[11]

In 2023, a judge of the High Court of Punjab and Haryana used ChatGPT to gauge the scope of bail jurisprudence.[12] The applicant, who was remanded in custody for his involvement in a brutal assault which caused the death of another person, applied for bail. In deciding to refuse bail, the judge reasoned that "[w]hen the physical assault is done with an element of cruelty, the parameters of bail also change". His Honour went on to say that "[o]nce the courts form a prima facie opinion that the accused acted with cruelty, then such an accused ordinarily should not be granted bail".

In the judge's "post-reasoning", his Honour stated that "[t]o further assess the worldwide view on bail when the assault was laced with cruelty, the use of an Artificial Intelligence platform which has been trained with multitudinous data was made". The judge asked ChatGPT "what is jurisprudence on bail when the assailants assaulted with cruelty?". ChatGPT's response was included, in full, in the judgment.

Despite that apparent substantive use of the AI platform, the judge did note that "any reference to ChatGPT… is neither an expression of opinion on the merits of the case nor shall the trial Court advert to these comments. This reference is only intended to present a broader picture on bail jurisprudence, where cruelty is factor".

Judges of the Bolivian Constitutional Court apparently consulted ChatGPT during an online hearing involving three journalists accused of publishing photos of a victim of violence without her consent. A news report about the case records that the judges asked ChatGPT if there was any "legitimate public interest" in journalists posting such photos without consent. ChatGPT responded that it was a "violation of the person's privacy and dignity". The Court ultimately ordered the removal of the photos. Although the Court mentioned that ChatGPT only helped in "clarifying certain concepts", the journalists' lawyer is reported to have described its use during the hearing as "arbitrary" and a "disaster", saying that "it can't be used as if it's a calculator that takes away the obligation of judges to use reason and to apply justice and to apply it correctly". The lawyer also observed, astutely, that "ChatGPT doesn't stop being a robot. If you ask it in the right way, it will answer what you want to hear".[13]

I have not used it in the course of my work yet. As the tools available to us improve, in terms of the information or data on which a generative AI chatbot is trained, and our knowledge and skills improve, I am sure we will find appropriate ways to use this technology in a way that improves efficiency and

---

11 There is a discussion of this case in an article in The Guardian online newspaper.

12 *Singh v State of Punjab* (CRM-M-22496-2022).

13 A Smith, A Moloney and A Asher-Schapiro, "Are AI chatbots in courts putting justice at risk?", *Context* (website, 04 May 2023) < https://www.context.news/ai/are-ai-chatbots-in-courts-putting-justice-at-risk>.

saves time, without detracting in any way from the fundamental principles of judicial independence and the rule of law: citizens engaged in a legal dispute are entitled to and should expect an independent judicial officer to bring their human mind to the resolution of that dispute, consistent with the application of the rule of law and all that entails.

Turning then to the next challenge….

2. Bias: The Ghost in the Machine

As CoPilot observes, "GenAI, despite its silicon neutrality, inherits biases from its data lineage. It mirrors the world it was trained on – flaws and all. Imagine an insurance claim assessed by an AI that unwittingly favours certain demographics or perpetuates systemic inequalities. As guardians of justice, we must exorcise these biases, lest our legal system becomes a rigged game." That's a fair statement.

As a powerful demonstration of the potential for bias, I asked CoPilot if it could please create me an image of a Chief Justice in Queensland in 2024. It took some time to think about it – making me feel as though I was engaging with a human being, as it is designed to – before producing this:

Say no more.

But more broadly, bias, like hallucinations, is one of the, arguably, inevitable limitations of a tool created using machine learning. Because the output is predictive, based on training data, that output may be skewed in some way, depending upon what the training data is (let alone the prompt asked, or the probabilities engaged). It is important to be aware of this possibility and the need to correct it.

The Australasian Institute of Judicial Administration (AIJA) has produced a detailed report in relation to AI Decision Making and the Courts. Among a number of other topics, it addresses the question of bias and discrimination in AI tools, and makes the point that "bias in human systems can be duplicated or enhanced in automated systems in different ways".[14] This includes the situation where the training data is not representative or is generated through biased human action and also potentially arises from humans over-relying, or putting too much weight on, AI systems, assuming they are "objective" or "scientific".

[14] AIJA's Report on AI Decision Making and the Courts (2022) at section 4.3.

As the authors of the AIJA report observe, “principles of impartiality and equality before the law require not only that like cases are treated alike, but also that different cases are treated differently”. The limitations of machine learning are such that AI systems are unlikely to be compatible with these two fundamental principles, at least in so far as substantive analysis and decision-making is concerned.

3. Copyright Quandaries: and Confidentiality

The next ethical quandary or challenge identified by CoPilot itself is copyright. I would add under this heading the issue of “confidentiality”.

As to copyright, the guidelines issued for non-lawyers in Queensland include the following caution:

> “The use of AI tools based on LLMs may also raise copyright and plagiarism issues. For example, Generative AI chatbots can be very useful in condensing or summarising information or presenting the information in a different format. However, the following should be considered:
>
> - using a 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.”

It is essential to consider confidentiality when entering any information into a generative AI chatbot – depending on the platform you are using, any information entered could become publicly available, as part of the “training data”, as could your “prompts”. It is a matter of being aware of how particular platforms work, and whether the data that you wish to upload does become available for “scraping” more broadly, or not, and whether your prompts are saved, or not.

4. The Fine Line: Human AI Collaboration

This was the fourth quandary identified by Microsoft CoPilot in response to my basic question. In so far as substantive decisions are concerned, I don’t think the line is fine at all. I think it is a hard one, that we should not cross. We should use available technology where it saves time and money for clients and, in my case, litigants (summarising information and legal research once those platforms become available, are good examples). But we are part of a human system, which requires human decision-making, with all that that involves, including the ability to reason, draw inferences, think critically and act compassionately.

5. The Regulatory Maze: Navigating Uncertainty

So far, the regulation of the use of AI in the context of courts has been by the publication of guidelines, either by courts or professional associations. New Zealand were the front-runners, leading the way with publication of guidelines for judicial officers, lawyers and non-lawyers. They were followed by the Courts and Tribunals of the UK.

I have referred already to the guidelines issued by Queensland Courts and Tribunals for Non-Lawyers. We also have internal guidelines for judicial officers. We did not publish guidelines for lawyers, because we took the view that the professional bodies, the QLS and BAQ for example, were better placed to do that, given the broad considerations to be taken into account. The Queensland Law Society has since published a Guidance Statement on the use of Artificial Intelligence in Legal Practice, and I understand the Bar Association of Queensland is currently considering such guidelines for its members.

The Supreme Court of Victoria has also published guidelines, as have the New South Wales Bar Association and Law Society.

As Sir Geoffrey Vos said of the UK guidelines – which applies just as much to the Queensland guidelines:

> "The messages contained in the Judicial Guidance are very simple. They apply just as much to lawyers as well.
>
> They can be summarised as follows.
>
> First, before using generative AI, you need to understand what it does and what it does not do. Generative AI does not generally provide completely reliable information, because the LLM is trained to predict the most likely combination of words from a mass of data. It does not check its responses by reference to an authoritative database. So, be aware that what you get out of an LLM may be inaccurate, incomplete, misleading or biased. …
>
> Secondly, lawyers and judges must not feed confidential information into public LLMs, because when they do, that information becomes theoretically available to all the world. Some LLMs claim to be confidential, and some can check their work output against accredited databases, but you always need to be absolutely sure that confidentiality is assured.
>
> Thirdly, when you do use a LLM to summarise information or to draft something or for any other purpose, you must check the responses yourself before using them for any purpose. In a few words, **you** are responsible for your work product, not ChatGPT."

More broadly, regulation flows form the general law (for example, as to copyright or confidentiality) and from the professional conduct rules that bind solicitors and barristers.

An open question is whether there is a need to disclose that you have used generative AI – at present, there is no such obligation under Queensland's guidelines.

Similarly, there is presently no obligation to disclose the use of AI in the UK. However, the UK Bar Council's guidance paper states that "[b]arristers 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".

In New Zealand, the guidelines provide that lawyers do not need to disclose use of a generative AI chatbot as a matter of course, unless asked by the court or tribunal.

In Victoria, the Supreme Court's guidelines state that ordinarily, practitioners should disclose to each other the assistance provided by AI programs to the legal task undertaken and that where appropriate (for example, where it is necessary to enable a proper understanding of the provenance of a document or the weight that can be placed upon its contents) the use of AI should be disclosed to the other parties and the court.

In contrast, in Dubai, parties are required to declare if they have used or intend to use AI-generated content during any part of proceedings. The guidance note provides that "[e]arly disclosure of the use or intention to use AI gives all parties the opportunity to raise any concerns they might have or to provide their consent to such use".

In Manitoba, the Court of King's Bench has issued a practice direction requiring that, where artificial intelligence has been used in the preparation of materials filed with the court, the materials must indicate how it was used.

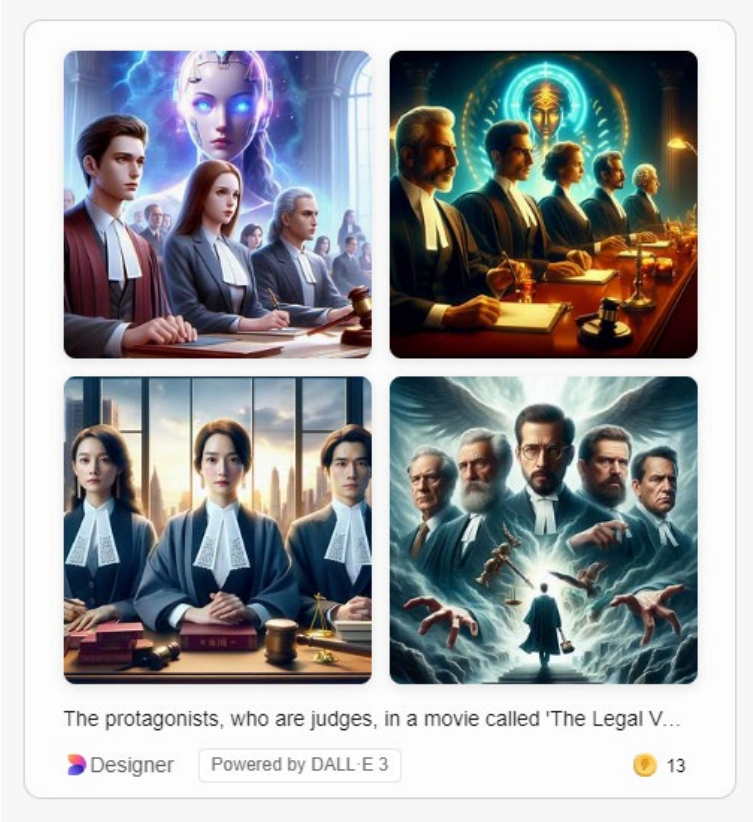
Even without an express obligation to disclose, lawyers, who are officers of the Court, are bound by their ethical and professional obligations, to ensure the accuracy of any document prepared to file, submit or tender in a court or tribunal.

This is a fast-changing topic, and it is to be expected that any published guidelines will be regularly reviewed, revised and updated. Issues such as whether disclosure is required as a matter of course will evolve, as experience of the use of AI in court proceedings grows.

I will let Microsoft Copilot have the last word:

> "In this twilight zone of GenAI, we must wield our legal wands judiciously. Let us embrace the magic, but with eyes wide open. For the ethical compass that guides us transcends algorithms – it pulses within our hearts, reminding us that justice, even in the age of AI, remains a human endeavour.
>
> So, my fellow legal voyagers, let us navigate this brave new world – one algorithm at a time."

That sounds like the opening narration of a very bad movie; perhaps it might be called “The Legal Voyage – Venturing into the Abyss of AI”. As to who the protagonists of this epic adventure might be, CoPilot has some suggestions:



Thank you.