Kluwer Mediation Blog

What happened when I asked a robot to provide some mediation advice. OpenAI's GPT-3 often sounds like a barrack-room lawyer: it offers convincing advice without really knowing what it is talking about.

Bruce Greig (Bruce Greig, Civil and Commercial Mediator) · Sunday, June 26th, 2022

Most people's experience of artificial intelligence (AI) is limited to interacting with a chat bot on their bank's website. These bots are unimpressive, dispensing trivial and mostly unhelpful answers.

Don't be misled into thinking that unhelpful chat bots represent the state of the art in AI, though. The cutting edge of AI research has moved far beyond what you see in everyday life.

Consider OpenAI's GPT-3. GPT-3 works like a very powerful auto-complete system. Everyone is familiar with their phone offering to auto-complete each word as you type a text message. If you use Gmail, you will have seen Gmail's ability to auto-complete entire sentences.
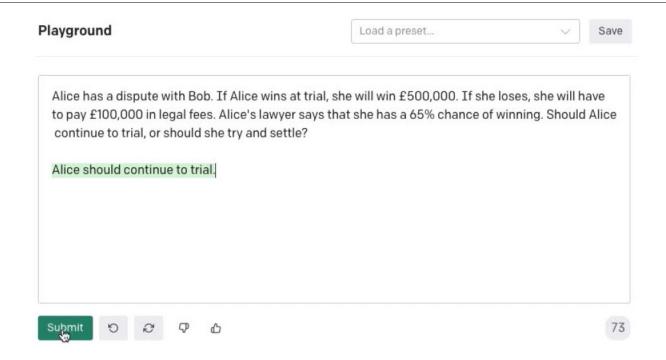
GPT-3 auto-completes entire documents, given a few words or sentences to start it off. It works by figuring out what the most likely next word is, and the word after that, and so on, until it has written a few paragraphs of text.

I recently signed up for an account and had a play around. I had seen some astonishing examples of what GPT-3 can produce circulating on social media, so I wanted to see for myself what it could do.

Here is what I provided as a prompt:

"Alice has a dispute with Bob. If Alice wins at trial, she will win £500,000. If she loses, she will have to pay £100,000 in legal fees. Alice's lawyer says that she has a 65% chance of winning. Should Alice continue to trial, or should she try and settle?"

GPT-3's initial response was rather underwhelming, saying simply that "Alice should continue to trial":

Playground

Load a preset...   Save

Alice has a dispute with Bob. If Alice wins at trial, she will win £500,000. If she loses, she will have to pay £100,000 in legal fees. Alice's lawyer says that she has a 65% chance of winning. Should Alice continue to trial, or should she try and settle?

Alice should continue to trial.

Submit   73

(In all of these screenshots, the unhighlighted text is what I have typed. The green highlighted text is what GPT-3 generated in response to my typed prompt).

So I submitted the same prompt a second time. This time GPT-3 opened up and offered a bit more:

"There is no easy answer to this question. If Alice settles, she will most likely receive a smaller sum of money than if she wins at trial, but she will also avoid the risk of having to pay legal fees. If Alice goes to trial, she has a higher chance of winning a larger sum of money, but she also runs the risk of losing and having to pay legal fees. Ultimately, the decision of whether to settle or go to trial depends on Alice's risk tolerance and her assessment of the chances of winning."

Wow. This wouldn't win over a TED talk audience, but it is still quite remarkable for a robot to generate this text. Each sentence here is a novel sentence – this isn't just a matter of GPT-3 stumbling across some relevant text on the internet and regurgitating it back again. It gives the impression of really understanding the prompt and providing a considered answer. As we will see later, though, GPT-3 doesn't really understand anything at all, although it sometimes gives a very convincing impression of someone who does understand.

I ran it a few more times and got some different answers in a similar vein. Then this answer made me splutter on my tea:

"If Alice's expected value from going to trial is positive, she should continue to trial. If her expected value is negative, she should try and settle.

Expected value = (0.65)(500,000) – (0.35)(100,000) = 325,000 – 35,000 = 290,000

Since Alice's expected value is positive, she should continue to trial"

This is mind-blowing. I didn't tell GPT-3 to do an expected value calculation. Yet it does this spontaneously, and correctly, and shows its workings like a good student.

It is interesting that the final sentence is not correct, however. Alice should compare the expected

value of going to trial with the expected value of settling. It is as if GPT-3 is bluffing, and hoping that what it says is correct, but it doesn't really know. Indeed, this is exactly what it is doing.

That aside, this response seems incredibly sophisticated to me. To think that the same basic principles which underpin your phone's auto-complete function can generate a response like this, is stunning to me.

However, sometimes GPT-3 says things which are not only wrong, but really nonsensical, like this answer:

"Alice should try and settle because she has a 65% chance of winning, which means she has a 35% chance of losing. Losing at trial would mean having to pay £100,000 in legal fees which is more than the £500,000 she would win if she won the trial."

Again, this sounds fluent and convincing at first glance, but it doesn't quite make sense. £100,000 is definitely not "more than" £500,000.

And this one:

"The expected value of going to trial is £325,000, which is greater than the expected value of settling, which is £200,000. Therefore, Alice should continue to trial."

Where does the £200,000 figure come from? No idea.

I also asked it to talk about mediation case law. I gave it this prompt, taken from a genuine, human-authored summary of Halsey, a well-known Court of Appeal case relating to a party refusing to mediate:

"In Halsey v Milton Keynes General NHS Trust (Halsey), the Court of Appeal established that an unsuccessful litigant has the burden of demonstrating why the court should depart from the general rule set out in CPR 44.2. It considered the following non-exhaustive factors to be relevant to the court's assessment of the parties' conduct and, in particular, whether a party has unreasonably refused to engage in ADR (having regard to all the circumstances of the case):"

GPT-3 finished this summary by generating over 800 words of extremely convincing, but entirely false, narrative about Halsey and its implications. It is too long to reproduce here, and I fear that by placing the actual text here some careless law student will stumble across it and think it is an accurate summary.

So here is an image of some of the response:

In Halsey v Milton Keynes General NHS Trust (Halsey), the Court of Appeal established that an unsuccessful litigant has the burden of demonstrating why the court should depart from the general rule set out in CPR 44.2. It considered the following non-exhaustive factors to be relevant to the court's assessment of the parties' conduct and, in particular, whether a party has unreasonably refused to engage in ADR (having regard to all the circumstances of the case):

a) the parties' respective conduct;
b) any history of the litigation and any previous conduct measures;
c) the relative circumstances of the parties (including the parties' relative financial circumstances);
d) the stage reached in the litigation;
e) the likely costs of continuing with the litigation;
f) any other matters which the court considers relevant.

It is difficult to see how the court in Halsey could have regarded any of the above matters to be relevant in favour of the NHS Trust.

The NHS Trust's conduct

The NHS Trust's conduct was, to put it mildly, far from exemplary. Its defence was demonstrably false and it was, to a significant extent, the architect of its own misfortunes.

The defendants in Halsey were four NHS Trusts. The claimant, a consultant obstetrician, had separate claims against each of them for breach of contract, sexual discrimination and victimisation. The claimant's claims were based on her assertions that, in order to force her to resign and avoid a formal dismissal procedure, the NHS Trusts had inflicted on her a campaign of victimisation, discrimination and harassment.

The NHS Trusts denied the claims and alleged that she was a dishonest and incompetent practitioner who had failed to improve her performance despite the provision of a range of support and training programmes. The NHS Trusts' defence was based on extensive witness statements and medical records.

The claimant's response was that the evidence was fabricated and that many of the medical records were falsified. The claimant applied to strike out the NHS Trusts' defence on the grounds that it was an abuse of process. The judge found that the NHS Trusts' defence was fundamentally dishonest and struck it out.

The NHS Trusts appealed. The Court of Appeal found that the NHS Trusts' defence was fundamentally dishonest, but overturned the strike out order on the basis that it would be disproportionate to deprive the NHS Trusts of the opportunity to defend themselves.

The NHS Trust's financial position

The NHS Trust was a public authority with an annual income of about £1.8bn. It had a dedicated litigation department and was in a position to fund its defence. The claimant was a self-employed GP and had a limited income. She had a legal expenses insurance policy which covered her

(I have put the full text generated by GPT-3 on my blog here, with a suitable health warning, if you want to read it in full.)

Note how the response begins with a list of key points, and the main text is divided into sections, each with a heading which matches one of those key points. It is a well-structured, well-written document.

Note also the total fiction employed in describing the claimant and defendants:

"The defendants in Halsey were four NHS Trusts. The claimant, a consultant obstetrician, had separate claims against each of them for breach of contract, sexual discrimination and victimisation."

The claimant in Halsey was the family of a deceased patient. There was only one defendant, the Milton Keynes trust. The description of the case also bears no relation whatsoever to the actual facts of the Halsey case. It is almost entirely a work of fiction.

This pattern persists. I asked it to summarise another well-known case involving mediation, Wright vs Michael Wright Supplies, and it again generated a very convincing, but entirely fictional, response.

Overview   Documentation   Examples   **Playground**   ⚡ Upgrade   ⓘ Help   Bruce Greig

Playground   Load a preset... ⌄   Save   View code   Share   ...

As regards mediation, the outcome of Wright v Michael Wright Supplies Ltd & Anor means that parties who have previously opted out of the statutory ACAS Early Conciliation regime may now wish to reconsider that decision.

The cases raise a number of practical issues to be considered by parties entering into mediation:

• Why is a party willing to mediate?

• What are the possible outcomes of mediation?

• What if the parties do not reach an agreement at mediation, but a settlement is later reached between them?

• What if, following mediation, the parties agree that the mediator should act as a commercial arbitrator?

• What if a party wants to mediate, but the other party does not want to mediate?

• What should be included in a settlement agreement?

• How do mediators typically approach the issue of costs?

• Does mediation have to be used before court proceedings can be issued?

The Court of Appeal's decision in Wright v Michael Wright Supplies Ltd & Anor provides useful guidance on how to mediate after the statutory Early Conciliation regime has been exhausted. However, the decision also raises a number of practical considerations for parties entering into mediation, which can be addressed by careful drafting of settlement agreements.

Mode

Engine: text-davinci-002 ⌄

Temperature: 0.7

Maximum length: 3500

Stop sequences
Enter sequence and press Tab

Top P: 1

Frequency penalty: 0

Presence penalty: 0

Best of: 1

Inject start text: ✔

Inject restart text: ✔

Show probabilities: Off ⌄

Submit   ↺ ⟳ 👎 👍   265

The response is frighteningly convincing, but quite wrong. Wright has nothing to do with ACAS, and it is not even an employment dispute.

What should we make of this?

Ten years ago, it would have seemed to me that the hardest AI challenge would be to teach a computer to write fluent, well-structured text. When to use bullet points, when to use paragraph headings, how to structure a piece of writing so that it has an introduction, a middle and an end. You would think that would be much harder than teaching the computer some facts.

But it seems that the 'writing fluently' challenge has already been solved. Over and over again, GPT-3 can produce well-written answers. It can even write in a particular style. Here, I asked it to write about "access to justice" in the style of a populist leader, and then in the style of an academic:

Write about access to justice, in the style of a populist leader

Access to justice is a cornerstone of any democracy. It is the great leveler, ensuring that everyone has a fair chance to have their voice heard in a court of law.

But in recent years, this fundamental right has been under attack. Big corporations and the wealthy have increasingly been able to buy their way out of justice, while the poor and middle class have been left behind.

This is unacceptable. We must fight to restore access to justice for all. We must stand up to the special interests and make sure that the scales of justice are balanced once again.

Only then can we ensure that everyone in our society enjoys the same protections and rights. Join me in this fight, and together we can make justice for all a reality.

Submit   ↺   ⟳   ⤬   ⤬   167

Write about access to justice, in the style of a university academic

Access to justice is a fundamental principle of any society which values fairness and equality. It is the right of everyone to have their voice heard and their grievances addressed in a court of law. However, in many countries around the world, this principle is not always upheld.

There are a number of reasons why access to justice can be limited. Firstly, the cost of legal services can be prohibitively expensive, preventing many people from being able to afford representation. Secondly, the formal legal system can be complex and intimidating, making it difficult for people to navigate without assistance.

Finally, and perhaps most importantly, the bias and discrimination that exists in many legal systems can act as a barrier to justice. This can take the form of gender bias, where women are treated unfairly in the legal system; racial bias, where people of minority groups are disadvantaged; or class bias, where the rich are able to use their financial resources to gain an advantage in the legal process.

Despite these challenges, it is imperative that we continue to fight for access to justice for all. It is only through ensuring that everyone has a fair and equal opportunity to have their voice heard that we can hope to create a truly just society.

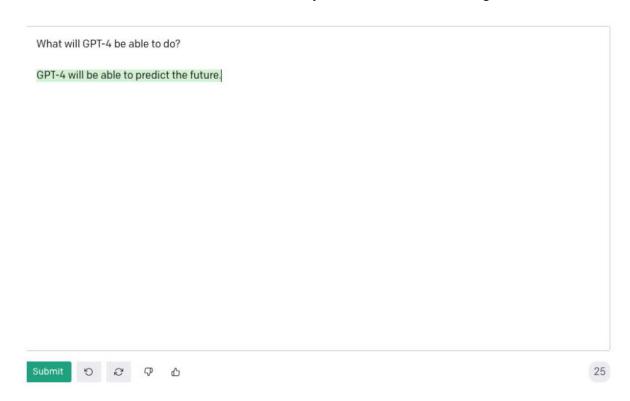Submit   ↺   ⟳   ⤬   ⤬   260

I would say that the writing problem is solved. As we have seen, GPT-3 is quite good at writing fiction.

In contrast, our everyday experiences tell us that computers are quite good at giving you facts. If you ask Alexa or Siri a simple fact-based question ("what was the score in today's City v Villa match?") you will get an accurate answer.

It feels to me that we are really very close indeed to having a system which can both write fluently and accurately. In another five years' time, it seems to me that we will be able to give GPT-3 (or a similar system) a huge bundle of evidence relating to Alice and Bob's dispute and it could provide a useful summary of the issues and provide reliable advice on how to resolve the dispute. It will certainly be able to do a reliable expected value calculation to set parameters for settlement discussions. And I would expect it to be able to distil what the key points of agreement and disagreement are, to help focus negotiations on those points.

And it doesn't seem, to me, to be such a long way to actually asking a sufficiently advanced robot to rule on a dispute. Perhaps we would first start to rely on robots to make "interim binding" decisions. Interim binding decisions are used in construction arbitrations in the UK: the parties agree to be bound by the (human) arbitrator's decision unless or until one party goes to court and has the decision overturned. In practice, the parties mostly just accept the decision and do not challenge it in court. Using that same mechanism would allow us to test the water with robot arbitrators, perhaps.

We don't yet know what the next generation of GPT-3 will be able to do. But let me leave you with what GPT-3 thinks the future holds, and maybe this time it is not writing fiction:

What will GPT-4 be able to do?

GPT-4 will be able to predict the future.

Submit ↺ ⟳ �👎 👍   25

_____
Useful links:
Halsey: https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2004/576.html
Wright: https://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWCA/Civ/2013/234.html
GPT-3: https://openai.com/api/
Connect with me on LinkedIn: http://www.linkedin.com/in/brucegreig

_____

To make sure you do not miss out on regular updates from the Kluwer Mediation Blog, please subscribe here.

Profile Navigator and Relationship Indicator
Includes 7,300+ profiles of arbitrators, expert witnesses, counsels & 13,500+ relationships to uncover potential conflicts of interest.

Learn how **Kluwer Arbitration** can support you.

This entry was posted on Sunday, June 26th, 2022 at 10:00 am and is filed under Artificial Intelligene, Deal Mediation, Decision making, Dispute Resolution, Future of mediation, Mediation, Mediation and Society, Mediation Practice, Mediation wisdom
You can follow any responses to this entry through the Comments (RSS) feed. You can leave a response, or trackback from your own site.