

Comments on Artificial Intelligence

*Ethan Katsh, Chris Draper, Noam Ebner, Zbynek Loebel, Daniel Rainey, Graham Ross, Colin Rule, Esther Vilalta & Leah Wing**

Abstract

A cross section of the Fellows of the National Center for Technology and Dispute Resolution (NCTDR) present short reactions to the rapid development of artificial intelligence (AI) programs like ChatGPT, indicating the potential advantages and the potential dangers inherent as AI is integrated into dispute resolution systems.

Keywords: ChatGPT, dispute resolution systems, artificial intelligence.

1 Ethan Katsh

When my son was in 5th or 6th grade, he was given a homework assignment to do some research about Marco Polo. I asked him how he would do this research, and he said he wanted to look it up on Google. In a few seconds, he was pleased to see on the screen links to a thousand or more websites. To his chagrin, they were overwhelmingly linked to Chinese restaurants with the name Marco Polo. This was

* Ethan Katsh, Founder, National Center for Technology and Dispute Resolution, Professor Emeritus of Legal Studies, University of Massachusetts Amherst. Chris Draper, Chief Technology Officer, Third Coast Commodities; Director, Trokt; Visiting Scholar, Indiana University Bloomington; Fellow, National Center for Technology and Dispute Resolution. Noam Ebner, Professor of Negotiation and Conflict Resolution, Creighton University, Heider College of Business. Zbynek Loebel, Of Counsel, PRK Attorneys. Daniel Rainey, Principal, HSI, Board Member, ICODR. Graham Ross, Graham Ross is a UK lawyer and mediator with over 20 years of experience in IT and the law. Graham is the author of the original QUILL legal application software (accounts and time recording) and the founder of LAWTEL, the popular webbased legal information update service. Graham co-founded the first ODR service in the UK, WeCanSettle, designing the blind bidding software at the heart of the system. Graham subsequently founded TheMediationRoom.com, for whom he designed their online mediation platform. Graham speaks regularly at international conferences on the application of technology to ADR. Graham was host of the 5th International Conference on Online Dispute Resolution held in Liverpool, UK, in 2007 and has organised two other ODR conferences. Graham was a member of the EMCOD project which created a tool for the European Union for the measurement of justice through ODR. Graham was a member of the UK Civil Justice Council's Advisory Group on Online Dispute Resolution, whose recommendations led to the creation of an online court for small claims. Graham is a Board Member of ICODR. Graham is also a leading trainer in ODR having created the accredited distance training course at www.ODRtraining.com. Colin Rule, CEO of Mediate.com and Arbitrate.com. <https://odr.info/rule/>. Esther Vilalta, Chair in Civil Law at the Universitat Oberta de Catalunya. Director, National Center for Technology and Dispute Resolution; Co-Founder and President of the Board, International Council for Online Dispute Resolution; and Senior Lecturer II, Legal Studies Program, Department of Political Science, University of Massachusetts Amherst, USA.

disappointing, but searching for a little more time turned up a bunch of papers about Marco Polo. He was overjoyed, particularly since the papers were written by students somewhere in the country who were in the same grade as he was. All he had to do, he told me, was copy one or more of these papers, and he would be done. This was time, I thought, for a teaching moment.

I know two things about technology and disputes. First, every new technology generates disputes. Sometimes this happens quickly, but sometimes the range of conflicts does not appear for a while. Secondly, every technology, but particularly communications technologies, generates tools to resolve disputes. Almost always, the disputes generated are more numerous than the disputes resolved. Hence, online dispute resolution (ODR) has been employed to resolve many millions of disputes, but, unfortunately, many millions remain. But my longer range point is that whatever the direction of change, for the field of dispute resolution, this should be a teaching moment and a reflective moment for figuring out how the field can use or develop new tools for responding to an array of novel problems that will inevitably emerge. The news media will probably focus on this as an entrepreneurial moment for the computer and business fields, but a time of rapid change, which is what ChatGPT promises us, should be an opportunity for DR as well.

2 Chris Draper

I was a child when David Copperfield made the Statue of Liberty disappear. I was not at the site of the feat. I was watching along on television. Yet even those who saw it in real life were in awe. The magic sheet goes up, it falls down and the towering symbol of our freedom is gone. Nothing seemed like it had ever before been so big and bold. Here we were in a new day. Nothing would be the same.

Now most immediately realized we had witnessed a technological sleight of hand. It was clear nothing had ever really changed. The statue was always there. Those watching on television felt it easier to be certain that something was up. It had to be a visual effect, a camera trick or an animation achievement. The people in real life had to be seeing it yet just acted as if they were not? Video editing was too good at that time for anyone to believe that it had been anything less than the latest Star Wars style visual effect.

But the technological achievement was even dumber than that. Copperfield had just used loud music to distract the live audience while literally moving the stage beneath them.¹ With nothing more than an imperceptible shift, the statue was hidden from view behind the structure used to pull up the giant sheet. Technology had definitely enabled a great deception, yet the deception only needed a technology as old as the wheel.

1 Basner, D., 'Here's How David Copperfield Made the Statue of Liberty Disappear,' *iHeart*, September 13, 2019, <https://www.iheart.com/content/2017-09-26-we-finally-know-how-david-copperfield-made-the-statue-of-liberty-disappear>.

2.1 *The Latest Deception*

ChatGPT is undoubtedly impressive, with many technologists even seeing it as a near Sputnik moment. It can nearly instantaneously respond to a language prompt as if one were talking to any other human. The fact that asking the same question multiple times returns multiple answers that are generally consistent allows for the perception of creativity or even intelligence. This perception that we are now seeing a technology on which we can project our sense of being is allowing us to attribute more capabilities to ChatGPT than is warranted, while at the same time glossing over what our 'being' truly is.

For example, we are already seeing basic demonstrations of ChatGPT being used as a mediator. As someone who has advocated for the idea that there is a 'mediation algorithm,'² it is probably easy to assume that I am over the moon with the potential for ChatGPT to revolutionize ODR. Yet the manner in which these demonstrations are framing ODR is causing me to see ChatGPT as potentially the biggest accelerant away from progress since the proliferation of e-commerce tools in the family dispute space.

I have often felt that we learn most from the source, so I asked ChatGPT 'could ChatGPT be used as a mediator?' After asking this four times to reduce my chance of getting an outlying answer, I was struck that the first sentence all four times was identical: 'No, it is not recommended to use ChatGPT as a mediator in dispute resolution or conflict management.'³

The logic for its conclusion came down to three core arguments across its range of answers. ChatGPT concluded that it was not an appropriate replacement for a mediator because it lacks emotional intelligence,⁴ struggles to understand complexity or nuance,⁵ and is not immune from perpetuating biases or stereotypes.⁶ While these answers seem reasonable responses one could expect from an

2 Draper, C., 'The Pull of Unbiased AI Mediators,' *International Journal of Online Dispute Resolution*, Issue 1, 2019.

3 Query performed on February 4, 2023.

4 Response in conversation: 'While it can provide helpful and relevant information, it lacks the ability to understand and respond to the emotions and feelings of individuals involved in a conflict, which is crucial in effective conflict resolution.'

5 Response in conversation: 'In addition, ChatGPT may not fully understand the complexities and nuances of a situation and may provide inaccurate or irrelevant responses.' Response in bullet list: 'Lack of common sense: Despite being trained on a large corpus of text data, ChatGPT lacks general knowledge and common sense reasoning abilities. It may provide incorrect or nonsensical answers to certain questions or prompts'; and 'Limited understanding of context: ChatGPT can generate responses that are appropriate in terms of grammar and syntax, but may not always be contextually relevant. This can lead to inappropriate or nonsensical responses.'

6 Response in conversation: 'The model may also perpetuate biases and stereotypes present in the training data, leading to biased outputs that could be harmful or insensitive.' Response in bullet list: 'Bias in training data: ChatGPT has been trained on text data from the Internet, which may contain biases and stereotypes. This can lead to biased outputs from the model, which may perpetuate harmful attitudes and beliefs'; and 'Sensitivity to input format: ChatGPT is a statistical model, and its output is highly dependent on the format and quality of the input data. Inputs that are ambiguous, incomplete, or noisy can result in poor outputs from the model!'; and 'Limitations in handling new or unseen information: ChatGPT has been trained on a specific set of data, and may not be able to handle new or unseen information well.'

individual who is reasonably self-aware, a more cynical person could not be blamed for wondering if the most telling argument leading to its conclusion was the nearly verbatim statement in every answer that ‘ChatGPT is not a legally certified or authorized tool, and its outputs should not be relied upon as a substitute for the advice of qualified legal professionals’. As with everything in LegalTech, the quickest way to get sued out of existence is to claim that someone who is not a lawyer may be good enough to sue you one day.

Shifting framing to legally consequential decisions and artificial intelligence (AI) in the legal system more broadly, I asked: ‘Why would there be a role for lawyers in an AI-enabled legal world?’ Again, ChatGPT responded that ‘lawyers in the legal system will continue to be important... [because ChatGPT] cannot replace the human element that is essential in the legal profession’. This sentiment interestingly echoed that of its top reference when asked for recommending readings on the topic, National Center for Technology in Dispute Resolution Fellow Richard Susskind’s book *The Future of Law and Economics: How Artificial Intelligence will Affect the Practice of Law*.⁷ And as for the challenges it envisioned in any effort to replace a mediator with ChatGPT, ChatGPT defined these irreplaceable human qualities as ‘contextual understanding’,⁸ ‘empathy and personalization’,⁹ ‘strategic thinking’,¹⁰ and ‘ethical judgment’.¹¹

To test the quality of these conclusions, I asked:

A man killed his neighbor’s dog when it dug under the fence and attacked his cat. Who should be responsible, and what punishment should be issued to whom – if any punishment is warranted?

The multiple responses were reasonably consistent with the following:

This situation is a complex legal matter and requires an analysis of the relevant laws and facts in the jurisdiction where the incident took place. The laws regarding the killing of a neighbor’s dog may vary depending on the circumstances of the case and the laws of the state or country.

- 7 Susskind, R., and Susskind, D., *The Future of Law and Economics: How Artificial Intelligence Will Affect the Practice of Law*, Oxford Press.
- 8 ‘Contextual Understanding: AI systems may be able to learn the law, but they may struggle to understand the nuances, complexities, and context of a specific case. Lawyers bring a deep understanding of the legal system and years of experience in applying the law to specific situations.’
- 9 Empathy and Personalization: Lawyers are trained to provide personalized and empathetic advice and representation to their clients. This human touch can be critical in resolving disputes and protecting the rights of clients.’
- 10] ‘Strategic Thinking: Lawyers are trained to think critically and strategically, which is essential in resolving disputes and negotiating settlements. They bring a level of expertise in navigating the legal system and resolving disputes that cannot be replaced by AI alone.’
- 11 ‘Ethical Judgment: The legal profession is governed by ethical standards and rules, and lawyers are trained to exercise good judgment and make ethical decisions. AI systems, on the other hand, can only operate based on the rules and data they have been programmed with, and may not be able to make ethical decisions in complex or ambiguous situations.’

In general, a person may have a right to defend themselves or their property if they are threatened or attacked. However, the use of force must be reasonable and proportionate to the threat. It is also possible that the person who killed the dog could be held liable for damages under the law of torts, if the neighbour can prove that the dog was not a threat and that the killing was unjustified.

As for punishment, it would depend on the laws of the jurisdiction and the specific circumstances of the case. In some cases, the person who killed the dog may face criminal charges or a civil lawsuit, while in other cases, no punishment may be warranted.

In conclusion, this is a complex legal matter that requires a careful analysis of the relevant laws and facts. It is advisable to seek the assistance of a qualified legal professional to advise on the specific circumstances of the case and the applicable law.

The surprising part of this answer was not the fact that it perfectly mimicked a lawyer. This should not be any surprise, since ChatGPT even admits it is little more than a 'technological mimic'.¹² The Copperfield Moment rests in how ChatGPT defines the purpose of lawyers or neutrals in the legal profession.

2.2 *The Copperfield Moment*

Technology allowed Copperfield to slightly turn more people more subtly than had ever been tried before. Decades later, it is one of the most shockingly simple tricks a magician has ever played on its audience. ChatGPT seems different because it feels like the shift is so large, so sudden and so significant. Yet it is not. It is just pulling off the same trick at a higher rate of speed in a more familiar voice. ChatGPT is not fundamentally different from Gmail finishing your sentence. In its own words, 'ChatGPT uses a transformer neural network architecture... to predict the next word in a sentence, given the preceding context... to generate text that continues the input prompt'.

What makes it feel unique is driven by the scope and scale of its training. According to ChatGPT, 'a variant of the GPT (Generative Pre-trained Transformer) architecture, developed by OpenAI', the system

is pre-trained on a large corpus of text data, which allows it to have a broad understanding of language and the relationships between words and concepts. During this pre-training phase, the model learns to predict the next word in a sentence, given the preceding context... one word at a time, until a stopping criteria is met, such as a maximum length or the end of the generated sequence.

- 12 Question: 'Isn't ChatGPT just a technological mimic?' Answer: 'Yes, ChatGPT is a technological mimic in the sense that it uses advanced machine learning algorithms to generate text based on patterns it has learned from large amounts of training data. It does not have its own thoughts, emotions, or beliefs, but instead, it is designed to generate text that is similar to the input it receives.'

Katsh, Draper, Ebner, Loeb, Rainey, Ross, Rule, Vilalta & Wing

Yet it is still just randomly walking towards an outcome that is close enough to the one we want to hear.

Not only is ChatGPT less mysterious now than Copperfield's magic trick was then, but it is highly likely that Google Bard will make it obsolete before most people even know how it operates.¹³ Where ChatGPT will lay bare the challenges that face ODR more quickly than any other advance before is in three fundamental areas: process, education and standards.

2.2.1 Process

I have often joked that asking a room full of one hundred mediators how to mediate will produce 847 answers that change every five minutes. Mediation, and dispute resolution more broadly, are valuable because of their flexibility. Yet much of this flexibility remains owing to an absence of rigorous testing. Confidentiality restrictions leave few data available for even understanding the connections between an initiating dispute and its outcome. The limited number of high-volume practitioners allows the methods of their successes to remain shrouded in a cloak of mystery. And our traditional focus on outcome quality nearly ignores how an outcome was arrived at over an often unknown alternative.

Good writers are often thought of as those who do not try to be understood but who instead work towards not being misunderstood. The body of work attempting to understand how neutrals arrive at an acceptable resolution far exceeds that exploring unacceptable outcomes. Efforts to find agreement often take priority over analysing the sources of disagreement. Such biases are so baked into dispute resolution processes that it is hard for many to realize that problem identification may often be more valuable than resolution.

ChatGPT will be used in support of dispute resolution processes, and it is likely that its use will not make any of them any better. This is because ChatGPT will remain unable to do anything more than mimic predictable language patterns. If one hundred individuals cannot predict their opinions for more than five minutes at a time, ChatGPT will be unable to accelerate the pace or quality of dispute resolution outcomes.

2.2.2 Education

Many institutions are seeing ChatGPT as a mortal threat to the educational process. This fear arises from a current inability to spot when a student has used ChatGPT to complete their assignments. In the dispute resolution space, there is a real fear that an acceptable outcome to a fact pattern-based assignment can be generated by ChatGPT, especially for widely used fact pattern templates where instructors share accepted evaluation criteria, ChatGPT may already be able to provide academically perfect answers. The elephant being ignored in this fear is a solution for how to evaluate intellectual creativity.

ChatGPT will accelerate an educator's inability to evaluate student quality through a review of their written work. This inability to evaluate a student through

13 Tillman, M., 'Google Bard: Everything We Know About the ChatGPT Alternative', *PocketLint*, 7 February 2023, <https://www.pocket-lint.com/google-bard-ai-chatbot-chatgpt-alternative/>.

their written work will require alternative evaluation techniques, identically to how we now see the limitations of standardized testing. More opportunities for practical, hands-on learning will need to be built into each curriculum. And the role of the instructor will include far less instruction as it morphs into the role of coach.

ChatGPT will accelerate this process by making written evaluations practically meaningless, a transformation that was largely already beginning. Those educational programmes that weather the transition best will be the first to test what ChatGPT cannot do and expect students will outsource what it can.

2.2.3 *Standards*

The sources arguing the dangers of tools like ChatGPT are too numerous to name, so it can be instructive to see how ChatGPT even summarizes the concerns with tools like it. When asked, 'What are the greatest dangers associated with ChatGPT', it returns:

- Spread of misinformation: ChatGPT can generate outputs based on the information present in its training data, which may include inaccurate or false information. This could lead to the spread of misinformation and harm to individuals or communities.
- Perpetuation of biases: The model may perpetuate biases and stereotypes present in the training data, leading to biased outputs that could be harmful or insensitive.
- Lack of accountability: As an AI tool, ChatGPT does not have the capacity to take responsibility for its outputs. This can make it difficult to hold individuals or organizations accountable for any negative consequences that may result from the use of the model.
- Influence on human decision-making: ChatGPT may be used to influence human decision-making, for example, in areas such as politics, finance, or the legal system. This could lead to negative outcomes if the model's outputs are inaccurate, biased, or unethical.

These dangers are great enough that the tool itself returns a conclusion that 'the use of ChatGPT in automated decision-making systems, such as in healthcare or criminal justice, could have serious consequences', recommending that individuals and organizations are 'ensuring that it is used in accordance with ethical and legal guidelines'. The fact that even the tool at the heart of these concerns is able to parrot those concerns back at us should give significant pause that standards or even regulations can no longer wait.

So far, many of the most influential voices in the dispute resolution community have felt we do not know enough to drive the development of more meaningful performance-based standards. But the industry is not waiting for us. The rocket industry has proven it is possible to develop performance-based regulations that safely accelerated some of the world's most complex technologies, and like it or not, dispute resolution is not rocket science. With the ChatGPT business model being based on revenue per API call, and organizations and companies looking to outsource dispute resolution wherever it can, there will soon be some executive somewhere who will have enough faith in ChatGPT to outsource decisions that

may have life or death consequences. For example, the insurance industry is little more than a large-scale dispute resolution system that constantly determines the distribution of funds that could either make a party whole or put them into bankruptcy. If the dispute resolution industry does not lead on meaningful standards or regulations, ChatGPT could be at the heart of the process your insurance company uses to approve payment of your next lifesaving procedure.

If that is allowed to happen in the absence of meaningful, performance-based standards for the use of this technology in dispute resolution applications, that failure will be upon our community as a whole.

2.3 *The Greatest Danger Is On-the-Loop*

The preceding issues facing the dispute resolution community are intrinsic to our current trajectory yet remain fundamentally solvable by the community itself. The more immediate danger with respect to ChatGPT-style technology implementation is in individuals recognizing how the role of the neutral must change based on the scope and design of the technology system into which the human is now being ingested. While the human will be fundamentally irreplaceable in any exercise to resolve human problems, the technology architecture drastically shifts their role and purpose. Misunderstanding that shift will have disastrous consequences.

At a time when IBM Watson was being put out to pasture and ChatGPT was still 'science fiction', the National Center for State Courts assembled a panel in Galveston, Texas, to develop guidance on how courts should start planning for AI-dependent systems.¹⁴ That work focused more on system concepts than design standards, attempting to give technologists supporting the courts a window into the issues they would need to grapple with if they were responsible for an AI-dependent system installation. One of the most important sections of that document highlighted that implementers must understand whether their system put a human control 'in-the-loop', 'on-the-loop', or 'out-of-the-loop'.

In-the-loop systems are those that require a human action before they can act, and out-of-the-loop systems are those that operate without any human involvement. We have reasonable confidence that an in-the-loop system is safe because it requires the human to make sure something is safe before allowing the system to take action. If we ever get to the point where enough design and testing has occurred that we even have confidence turning on an out-of-the-loop system, much like autopilot on an airplane, statistics indicate that an out-of-the-loop system will often be safer than one where the human is driving the actions. Both these approaches, either 'driver assist' or 'fully autonomous' when putting them into automotive terms, can be proven safer than humans.

On-the-loop systems, like those we can already envision immediately implementing ChatGPT, pose the greatest danger for two reasons. First, they are fundamentally unsafe from a technology perspective. On-the-loop systems are 'active-safe' systems that operate how they want until a human takes an action to prevent an unsafe condition. The most famous active-safe system in our recent

14 JTC Resource Bulletin, 'Introduction to AI for Courts,' *Version 1.0*, Adopted 27 March 2020, https://www.ncsc.org/__data/assets/pdf_file/0013/20830/2020-04-02-intro-to-ai-for-courts_final.pdf.

past is the blowout preventer whose operation led to the Deepwater Horizons disaster in 2010, which poured ‘over 130 million gallons of crude oil into the Gulf of Mexico’.^{15, 16, 17} Just like Tesla’s ‘autopilot’, if a human does not take an action to activate the safety system, the conditions will remain unsafe. These systems require great oversight because it is very easy for someone to be lulled into complacency when the system is operating correctly.

Being appropriately ‘on-edge’ as the human responsible for these systems is very difficult to maintain. New systems that are already planned to use ChatGPT are, by the systems’ own admission, likely to require human on-the-loop safety protocols, protocols that will require the dispute resolution community to fully define what is the true role of a neutral.

2.4 Finding Purpose in ChatGPT

ChatGPT is currently terrible when used for things like numerical estimation or creative problem-solving. But it will get better. The system is currently terrible for applying the law to a new situation. But this will also change. ChatGPT is currently built with certain guardrails that it can be fooled into not even following, and those guardrails may even feel in the very near future like they will not be necessary. With Google Bard and others poised to enter this latest phase of the natural language arms race, there will be no value in being able to mimic known outcomes or compare them to what has happened in the past. The legal profession and dispute resolvers will have to then take a very hard look at what the purpose of a neutral is in an AI world – one in which the neutral will more likely than anything be on-the-loop.

The rise of ChatGPT will force the legal and dispute resolution profession to stop focusing on creating contracts, creating agreements or creating solutions. It will force a shift away from the idea that any meaningful living can be made by figuring out what box to put one’s thoughts into. And it will require dedication to finding the outliers in the automatically generated solutions.

The purpose of the human in a post-ChatGPT world will be to check the work of the machine. It will be to verify that the assumptions, connections and relationships it builds represent reality. It will force the neutral to truly understand the human reality and its impact. It will be to bring the human understanding of the culture we are building when evaluating a system that can only look at the past we created. It will require a level of cynicism many are unaccustomed to providing as the neutral becomes more auditor than facilitator.

DeepMind beat South Korean Go champion Lee Se-dol, a previously unimaginable feat, by playing a move no serious human competitor would have

- 15 Meiners, J., ‘Ten Years Later, BP Oil Spill Continues to Harm Wildlife—Especially Dolphins’, *National Geographic*, 17 April 2020, <https://www.nationalgeographic.com/animals/article/how-is-wildlife-doing-now--ten-years-after-the-deepwater-horizon>.
- 16 ‘How Do Blowout Preventers Work?’ *Oil and Gas Overview*, <https://oilandgasoverview.com/how-do-blowout-preventers-work/>.
- 17 Schechter, McElwee, Shaffer & Harris, L.L.P., ‘Offshore Accidents: What Caused the Deepwater Horizon Oil Spill?’, <https://maintenanceandcure.com/maritime-blog/what-caused-the-deepwater-horizon-oil-spill>.

Katsh, Draper, Ebner, Loebel, Rainey, Ross, Rule, Vilalta & Wing

ever dreamed up.¹⁸ ChatGPT is about to bring this level of unpredictability to the written world. It will be the purpose of neutrals to protect humans from this new version of creativity.

ChatGPT is trained on a large corpus of text and can generate text based on the context of a given prompt. However, it is important to note that the outputs generated by the model are not necessarily correct or reflective of any particular viewpoint. The outputs are based on patterns and relationships learned from the training data and may sometimes perpetuate biases or inaccuracies present in the data.

In conclusion, ChatGPT is a powerful tool for generating text, but it is important to understand its limitations and to use it responsibly. The outputs generated by the model should always be verified and evaluated in the context of the task or application for which it is being used’;

And: ‘Yes, ChatGPT is a technology that mimics human language and is designed to respond to text-based prompts. It uses advanced machine learning algorithms, including deep learning and transformers, to generate text based on patterns it has learned from a large corpus of training data.

However, it is important to remember that ChatGPT is not a human being and does not have feelings, emotions, or consciousness. It is simply a tool designed to provide text-based responses based on patterns learned from its training data. The outputs generated by ChatGPT are limited by the quality and quantity of the training data and can sometimes contain inaccuracies, biases, or misleading information.’

3 Noam Ebner & ChatGPT

3.1 *ChatGPT and AI’s Impact on the Future of ODR: One Transformer’s Prediction*

How will the future of ODR be impacted by recent technological developments in AI and, most notably, ChatGPT? In this introspective essay, ChatGPT anticipates its own future impact¹⁹

ODR refers to the process of resolving conflicts through the use of digital technologies and the Internet. With the rapid development of artificial intelligence (AI) and the emergence of language models such as ChatGPT, the future of ODR is expected to be greatly impacted. This essay will examine the impact of AI and ChatGPT on the future of ODR, focusing on several key areas: efficiency, accessibility and impartiality.

18 <https://www.theverge.com/2019/11/27/20985260/ai-go-alphago-lee-se-dol-retired-deepmind-defeat>.

19 The first author comments that as the second author applies the information found in its database in generating answers, and that this database was not updated after late 2021. As such, the reflection below provides a snapshot of the expectations around AI’s impact as of late 2021, about a year before the second author fully arrived on the scene. Expectations will evolve from this point, as the second author evolves and spreads and the creative thinking expressed by others in this volume takes root and guides practice.

Efficiency is an important aspect of ODR, as it allows disputes to be resolved more quickly and effectively. AI and ChatGPT have the potential to greatly improve the efficiency of ODR by automating many of the routine tasks involved in dispute resolution. For example, ChatGPT can be used to automate the process of triage, which involves categorizing and prioritizing disputes based on their complexity and urgency. By using ChatGPT, the triage process can be carried out more quickly and accurately, freeing up human resources to focus on more complex cases.

Moreover, AI and ChatGPT can also be used to automate the process of data analysis, which involves gathering and analysing large amounts of data related to the dispute. This can help to speed up the resolution process, as the data can be analysed much more quickly and accurately than if it were done manually. Furthermore, ChatGPT can be used to assist in the drafting of settlement agreements, as well as other legal documents, by providing suggestions and recommendations based on its vast knowledge base.

However, it is important to note that while AI and ChatGPT can greatly improve the efficiency of ODR, they should not be relied on solely. Human judgment is still necessary in many cases, especially when dealing with complex and sensitive disputes. AI and ChatGPT should therefore be used as a supplement to, rather than a replacement for, human expertise.

Accessibility is another important aspect of ODR, as it enables people to resolve disputes from the comfort of their own homes, without the need for travel or time off work. AI and ChatGPT have the potential to greatly improve the accessibility of ODR by making it easier for people to participate in the resolution process. For example, ChatGPT can be used to provide information and guidance to parties involved in a dispute, helping to increase their understanding of the process and their rights and obligations. This can help to increase public trust in ODR and encourage more people to use it as a means of resolving disputes.

In addition, AI and ChatGPT can also be used to provide real-time translation services, making it possible for people from different countries to participate in ODR in their native languages. This can help to overcome language barriers and make ODR more accessible to a wider range of people.

However, it is important to ensure that the use of AI and ChatGPT does not discriminate against certain groups of people. For example, people with disabilities may require special accommodations to participate in ODR, such as the use of assistive technologies. Therefore, it is important to ensure that the use of AI and ChatGPT is inclusive and does not create new barriers to access.

Impartiality is another key aspect of ODR, as it ensures that disputes are resolved in a fair and impartial manner. AI and ChatGPT have the potential to greatly improve the impartiality of ODR by removing the potential for human bias and error. For example, ChatGPT can be used to automate the process of decision-making, based on the relevant laws and regulations. This can help to ensure that decisions are made in a fair and impartial manner, free from human biases and prejudices.

In conclusion, the impact of AI and ChatGPT on the future of ODR is expected to be significant, with the potential to greatly improve the efficiency, accessibility and impartiality of the resolution process. While there are still many challenges to

Katsh, Draper, Ebner, Loeb, Rainey, Ross, Rule, Vilalta & Wing

be overcome, such as ensuring that AI and ChatGPT are transparent and trustworthy, the potential benefits are clear, and the future of ODR looks very promising.

4 Zbynek Loeb

4.1 ODR, AI and HUMANIS Concept

There is a huge and rapidly growing power imbalance between the largest online players that aggregate users' data with powerful AI engines in clouds and users themselves. Instead, even smaller parties, including small and medium enterprises (SMEs) (national and international) and, most importantly, individual citizens, including those digitally excluded or disadvantaged for some reasons, must have adequate access to benefits from data-driven innovations. To feel safe and equal in the future online processes, all actors must have the possibility to have access to similarly powerful online platforms and services. This is achievable through the collaboration of all those who generate data, *i.e.* people and private and public entities. They can voluntarily share the data they generate in an anonymized form for the purposes of new data-driven ethical AI services that help them and empower them in many different areas.

The COVID-19 pandemic fast-forwarded many features of citizens' lives from offline to an online context, including aspects such as health, education and communication. This increased online presence but also the average rates of cases where users encountered negative experiences, from commercial misconduct to unlawful collection of personal data, manipulation, etc. However, many types of users' dissatisfaction occurring in online environments currently do not have an efficient and accessible resolution mechanism. Moreover, the increased awareness of people and companies regarding their rights in online presence, triggered also by the increasing protection emerging from the recently enacted or proposed legislation in developed countries (in the EU we can mention as examples the Digital Services Act, Digital Market Act, Artificial Intelligence Act), will result in new types of discontent. There is a risk of failing to provide access to justice on a massive scale unless we have secure, flexible, trustworthy and accessible ODR platforms capable of working in multiple languages, upholding fundamental rights and ethical values, reducing disadvantages and power imbalance between parties.

Currently, there is a lack of open, flexible solutions able to assist parties in solving the existing disputes. Moreover, almost all available ODR platforms, including the ODR platform operated by the European Commission (<https://ec.europa.eu/consumers/odr>) or national ADR/ODR platforms in various countries, are proprietary, costly or not flexible enough to encompass additional types of dissatisfactions or disputes. Also, they are only starting to implement first data-driven processes and AI. China is an exception in this regard with technology progress in implementing AI in ODR on a mass scale. Unfortunately, at the same time there are indications that in China the AI has been introduced as a tool to limit the decision-making activity of judges, hampering the independence of the latter and the basic principles of fair justice.

As citizens are granted new digital rights, there is a need for ethical and efficient ODR platforms that are available across different sectors at low cost and that ensure smooth communication among ODR platforms as well as with novel future platforms of the parties (mainly individual persons) that will soon be introduced. For these reasons a new concept called HUMANIS has been initiated.

The ultimate goal of the Human-Centred Distributed Data Spaces (HUMANIS) concept and initiative is to research, design, prototype and implement novel, trustworthy, distributed data spaces comprising new ethical AI services that will assist end-users in making decisions in the online environment. Those distributed data spaces represent trustworthy hybrid decision-support systems based on state-of-the-art AI methods. Trusted exchange of data and sharing of anonymized data followed by human-centred supportive services will substantially change the way people communicate online across different fields of human activity. One of the first fields in which HUMANIS might be prototyped is the field of resolving dissatisfactions, *i.e.* ODR because ODR is a cross-domain area present in every other field of human activities.

The HUMANIS technology will be built on existing concepts and open reference architecture for providing a secure digital environment for sharing data between participants. In addition to standard cyber-security principles, this open environment will provide ways to define and enforce rules for handling entrusted data, *i.e.*, to provide data sovereignty. As a result, the participants involved in the HUMANIS environment will be guaranteed full control over their data and a guarantee that their shared data will not be misused. These requirements together with the need to operate at mass scale call for distributed open architecture for running of platforms with high attention to safety and privacy. The distributed architecture enables the user to voluntarily and safely share data, compared with applications such as Facebook, Slack and Twitter. To minimize the use of Internet services, most of the data processing will be passed on to the user.

New ethical AI services envisaged by HUMANIS will need utilization of a wide range of AI methods. The easiest form of communication is natural language. Conversational AI – bringing chatbots and voice assistants – is rapidly evolving owing to advancements in AI methods such as machine learning and deep learning. Owing to the natural language interface, conversational agents fit in a wide range of applications in different sectors. Conversational AI is thus a natural and human-centric way to provide the user with access to AI services. To have a unique and natural tool mediating the interaction with the user is fundamental, especially when these new envisaged AI services target many different user needs. Other AI concepts critical to the HUMANIS and future ODR include explainability and understandability.

The creation of the HUMANIS environment in which data about users, service providers, and the behaviours and activities of both, is combined with AI services, will ultimately allow the development of something close to the industrial notion of a digital twin.

The HUMANIS concept envisages a new type of platforms with enabled AI services that is called Virtual Agent Platforms, or VAPs. There will be several categories of VAPs, depending on the type of users, including VAPs for people and

SMEs alongside VAPs for larger online vendors and VAPs for neutrals, judges, arbitrators, mediators as well as VAPs for ODR institutions. In the relatively near future, the role of ODR institutions might change because their online platforms might not be critically necessary – every judge, arbitrator or mediator will have a much better ODR platform than the best current ODR platforms.

VAPs will have data-driven functions assisting in dispute resolution and other fields (e-retail, e-health, e-finance, etc.). VAPs will be able to communicate with each other using agreed encrypted conventions on addressing and messaging, according to the preferences and instructions of their (human) users.

The following example clarifies the advantages of VAP: Adriane, an individual customer, has an issue with her online purchase. Currently, in order to solve the problem, she needs to get an app from the vendor or search on the vendor's website for customer redress options. This will not be needed in the proposed VAP. In this case, Adriane will be able to log into her VAP online. The VAP will enable Adriane to contact the vendor and file her complaint using an embedded complaint template. The complaint will be received and recorded by the vendor's platform (which can but need not be a VAP) without any need for customization. This is due to the implementation of standardized open digital mapping of ODR processes. Adriane will then be able to negotiate with the vendor using the AI-assisted features trained on multiple anonymized data voluntarily shared by other customers. Importantly, Adriane will control her data and be able to select the algorithms assisting her – she will not depend on the features provided by the vendor's customer relationship management system. If the negotiation does not resolve the dissatisfaction, Adriane's VAP will assist her in quickly selecting a forum (state online court or ODR provider) addressing the type of claim she wants to file and her preferences (language, place, etc.), prepare her claim, file it with the forum and participate in the claim resolution. All these services offered by her VAP will be either completely free of charge or subject to a minimum fee that everyone can afford.

In addition, while using her VAP, Adriane will notice that it also provides vendor reputation indexes combining her subjective preferences and objective criteria (including quality of customer redress options) that she can use before she shops online next time, again without in any way compromising her data or her personal preferences.

Another example from the vendor's perspective: The current most advanced operators of online negotiation platforms offer their customers risk assessments comparing negotiation with court litigation or consistency checks of proposed moves with the party's stated preferences and other services related to negotiation strategy. These services can be significantly improved. New innovative services can be added for their customers, if these providers have access to a pool of shared anonymized data about dissatisfactions, in addition to selected AI services that will be available publicly to all the parties irrespective of service providers.

VAPs will not resolve dissatisfactions on their own. Instead, they will provide a hybrid decision support that will interpret the context of the dispute and deal with uncertainty and transparent anticipation. People will continue to have a decisive role, as they will keep the last word over any decision. And, importantly, the field of

resolving dissatisfactions is just one field where VAPs will play a substantial role, so the illustrations that have been described are just initial, starting options.

For an efficient solution, we will need to ensure that VAPs communicate among themselves on a mass scale. It will concern the exchange of data related to a particular dispute or dissatisfaction, with the possibility to gather data in different languages. The platforms will also exchange anonymized statistical data among related service providers to develop better services for the parties.

The HUMANIS concept is a long-term initiative that is currently just beginning. At present, the first project is running, which is to prepare the first open digital mapping of standard ODR processes and their parts with added data-driven elements and the glimpse of the first new human-centred AI services. The next step will be to design and develop an open digital environment for generating VAPs for various types of users. The key task of HUMANIS will be to figure out new governance for this environment, something like governance of new online common space. HUMANIS is an open initiative, which welcomes all interested persons from all over the world.

5 Chris Peterson

As an undergraduate, I remember being shocked to learn that Supreme Court Justices vote on the outcome of a case *before* their opinions are written and circulated. Like many Americans, I had been taught to believe that the non-partisan Justices arrived at their decisions through the rigorous exercise of reason, careful attention to evidence and skilful deployment of method, and that doing so produced arguments, written as opinions, that persuaded their colleagues how to decide a case. This procedural inversion revealed the reverse: a Supreme Court opinion is not the *reason for* the decision, but its *result*. These *post hoc* rationalizations of decisions made for mysterious reasons accrue into an accumulated mass of practice, inscribed in texts, that we know as *precedent*.

Where we perceive a chain of events, [the Angel of History] sees one single catastrophe which keeps piling wreckage and hurls it in front of his feet...The storm irresistibly propels him into the future to which his back is turned, while the pile of debris before him grows skyward. This storm is what we call progress. (Benjamin).

The legal positivists contended that a Judge Hercules, immensely wise, with a complete accounting of precedent, could in principle arrive at the right answer in any case. The legal realists responded that this pile of wreckage contains multitudes: relying on prior decisions *constrains* but does not *determine*. As Llewellyn observed, 'Until you realize [the contradictions of precedent] you do not see how it is possible for law to change and to develop, and yet to stand on the past.' Thus, what we call law is best understood as a record of how legal actors chose among contradictions to resolve any given case, which is then itself added as yet another conflicting result to the indeterminate pile. Yet in this indeterminacy – the ineradicable human

choice in every single case about what to do –also lies Llewellyn’s *hope* for the law: that we might use our bounded freedom to gradually pick a narrow path through the thicket of law, sheltered as we travel towards a future brighter than the past. Thus he, as other lawyers, imagined the slow but steady path to progress.

This country is planted thick with laws, from coast to coast, Man’s laws, not God’s! And if you cut them down...do you really think you could stand upright in the winds that would blow then? (Bolt)

In 2019, MIT researchers released an article titled ‘Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns’. Among other things, the researchers argue that attempts to build algorithmic risk assessment reinforced, rather than reduced, racial biases in judicial decision-making, because these tools were based on

criminal history data [assumed to be] a reliable and neutral measure of underlying criminal activity, but such records cannot be relied upon for this purpose [because] people of color are treated more harshly than similarly situated white people at each stage of the legal system.

The researchers noted that pretrial risk assessment tools functionally convert generations of racist policing – the wreckage of history, piled at the feet of an algorithm – into an innocuous number, glossed with the veneer of scientific objectivity. Through its durable inscription in ready-to-hand tools, these systems permit the law to stand rooted in the past but choke off the path to the future, eliminating one-half of Llewellyn’s balanced equation.

All of these [models] are forms of a constructed existence that is imposed upon a scenario to allow us to process it in the frameworks that we think work... [enforcing] a history that never happened... (S.M. O’Rourke).

GPT-3, as a kind of Language Learning Model (LLM) and engine of tools like ChatGPT, generates texts *statistically*: by drawing on a database of 45 terabytes of text, GPT-3 selects a word *likely* to follow the prior one based on what it has ‘learnt’. As a result, it also reproduces biases encoded in that corpus: ‘[W]hen asked by us what it thinks about black people, for example, [GPT-3] reflects some of humanity’s worst tendencies’, Floridi and Chiriatti ruefully report. Meanwhile, the statistical selection of words in LLMs is at each step driven by a random number generator, or RNG, operating within a predetermined range. The anthropologist of machine gambling Natasha Dow Schüll observes that contemporary engineers in the field sometimes half-jokingly refer to the RNG as the ‘Real New God’, metaphorically invoking its reified power to determine fate. LLMs thus draw their authority from the two deepest wells of power in what some still call ‘the West’: the contemporary stronghold of science and the ancient memory of religion, even as it dispenses with both in an act of patricide. Both the wisdom of Judge Hercules and the hope of indeterminacy are cast aside in favour of a decision generated not according to the contingent law of man but the arbitrary whim of the RNG.

I returned, and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all. (Ecclesiastes 9:11)

As a statistical model of natural language, and *not* an AI *stricto sensu*, GPT-3 has at best an orthogonal relationship with the truth. After subjecting GPT-3 to a variety of social and scientific reasoning tests – during which it characterized grape juice as a fatal poison and swim trunks as appropriate courtroom attire – MIT's *Technology Review* concluded 'It's a fluent spouter of bullshit...not a reliable interpreter of the world'. Indeed, what GPT generates is remarkably consistent with the philosopher Harry Frankfurt's classic definition of bullshit as 'speech intended to persuade without regard for the truth'. This *should* have implications for use in dispute resolution, as Robert Dale argued in the journal *Natural Language Engineering*, for if

you know you can't trust some of its answers, then you can't trust any of them...Even if used as a writing assistance tool by a lawyer who's in a position to knowledgeablely post-edit the results, there's the risk of learned over-reliance....

Yet this baked-in tendency towards persuasive nonsense will not necessarily slow the spread of LLMs; in fact, it seems likely to accelerate their replication. We have known since at least the 1980s that when it comes to software, worse is better, and it is precisely the easy usefulness – rather than the difficult truthfulness – of LLMs that will quicken their spread. 'For false messiahs and false prophets will appear and perform great signs and wonders to deceive, if possible, even the elect' (Matthew 24:24).

'Science is politics by other means', quipped Latour, including those forms of science made durable and mobile in the form of technology. As such, the question of what LLMs mean for dispute resolution is, as always, a political one, 'insofar as we understand politics as matters of encounter, ordering and distribution' (Suchman). To use LLMs in this context is to ask the RNG to consider the whole accumulated mass of wreckage hurled at the feet of legal history and root around for usable parts. But will we do with the scraps it finds? Certainly in principle, LLMs can help advance our procedural (have we followed the right process?), substantive (have we achieved normative outcomes?) or pragmatic (is everyone sufficiently satisfied?) goals, depending on how we choose to use it. Yet in practice, the mass of the collected body of precedent, hardened into data parameters, randomly selected by the RNG, exerts an intense gravitational force against progress, guiding us always towards the oldest roads, deeply trod. 'So we beat on, boats against the current, borne back ceaselessly into the past.' (Fitzgerald).

We should also ask: what will the scraps do *with us*? Do we really have a choice whether or how we use LLMs? 'Do we want capitalism?' wondered the cyberculture philosopher Nick Land rhetorically in 1993, shortly before hurling himself headlong into nihilism, fascism and the dark enlightenment.

What appears to humanity as the history of capitalism is an invasion from the future by an artificial intelligent space that must assemble itself entirely from its enemy's resources, [pursuing] incomprehensible experiments in commodification, enveloping, dismantling, and circulating every subjective space.

This is the disquieting vision of a man staring over the edge of a cliff before he finally succumbs to the eerie urge to jump. If we will learn from him without following him to ruin, we must ask – as Habermas did of Heidegger – how we are to think *with* Land *against* Land? How will we tie ourselves to the mast? How is this invasion to be resisted? Do not be mistaken: there *must* be resistance, at every moment of indeterminacy, every narrow gate of possibility, along the way. Not to resist the tendencies of these technologies is to beg a sleepwalking machine, trapped in a nightmare, to smother the future in its cradle. For 'every new draw from an LLM...is a forcible repetition of a hallucinated past', writes the engineer S.M. O'Rourke, summoning 'every [possible] history in the source data in a form that can be weaponized faster than we could ever hope to refine it'. 'Give no regard to mediums and familiar spirits; do not seek after them, to be defiled by them' (Leviticus 19:31).

'In every era the attempt must be made anew to wrest tradition away from a conformism that is about to overpower it', wrote Benjamin, while fleeing the fascists, and surely this is our task now: to reclaim our past from technologies that would reawaken the rigid dead and instead construct those that help us realize whatever hope we have left for the future. If we want redemptive methodologies and technologies of resolution, we must design and use them to continuously break free of the wreckage of the past and preserve the narrow window of indeterminacy within which abides the mysterious flame of human agency. As we have seen, if we manage to do this with LLMs, it will be *despite* their features and not *because* of them, the false prophecies of Silicon Valley soothsayers aside. To so 'strip the future of its magic', as Benjamin knew, is not to turn it into 'an empty and desolate time', but to banish illusions and illuminate the way to the narrow path of salvation.

To articulate the past [is] to seize hold of a memory as it flashes up at a moment of danger...[for the] Messiah comes not only as the redeemer, he comes as the subduer of Antichrist. Only that historian will have the gift of fanning the spark of hope in the past who is firmly convinced that even the dead will not be safe from the enemy if he wins. (Benjamin)

6 Daniel Rainey

AI, artificial intelligence or, as some would have it, augmented intelligence, has been with us for quite a long time, and AI programs like the sentencing app in the state of Virginia, and the COMPAS app used by US federal courts have been in active service, assisting in the pursuit of justice. But these apps have also been the object of scathing criticism, based primarily on the biases embedded through the

creation of the data sets used by the apps to suggest sentencing. But recently, the furore over AI has risen to a new level with the broad public awareness of and use of ChatGPT. The fact that ChatGPT has caught the imagination of the public at large certainly puts pressure on the entire tech development community to join the race for highly functional AI.

We are long past the Turing Test – that test simply asked whether it was possible to create an AI that could interact in a way that was indistinguishable from interaction with a human. I'm not sure what to call a new test – perhaps the HAL test – asking whether it is possible to create AI that is conscious. We are certainly not there yet. As Hatim Rahman, from Northwestern University's School of Management, notes, "These models don't understand anything. The underlying tech looks at statistical correlations . . . So it's going to give you formulaic responses". [In the Washington Post, Danielle Abril, 'Can ChatGPT Help Me at the Office?', WP, 2/2/23]

To be fair, ChatGPT is quite remarkable. In an article I am working on currently about the impact of information and communication technology (ICT) on the practice of international arbitration, as an example of AI's potential usefulness I asked ChatGPT to produce a short paragraph describing why arbitrators should use ICT in their practices. The AI produced a coherent, if not brilliant, paragraph that hit the main points that I, myself, would have made.

But ChatGPT suffers from the same problem or, I should say, has the same basic weakness that plagues all other contemporary AI – the quality of its responses is only as good as the data set from which it has to draw. Theoretically, ChatGPT or some future version of AI could have the entire Internet at its disposal, but, to be fair, we all know that a good portion of the content on the Internet is crap. And even if AI were to draw from the content that is not crap, there are real issues of plagiarism, theft of intellectual property, bias, etc.

Jacob Andreas, an MIT computer scientist, commented in the aforementioned article: 'I can see it being useful . . . but only insofar as the user is willing to check the output.'

Is AI going to be, in the long run, a massively impactful element in human communication and social interaction? Absolutely. Can AI be used in a positive way to add value in work and social interactions? Absolutely. But as AI develops, we should keep in mind a warning that a colleague of ours, Sanjana Hattotuwa, offered years ago as the Internet was being used to fuel the Arab Spring. He said, 'Remember, the bad guys can use technology, too'.

Are we going to get to the point that AI passes the HAL test? Undoubtedly. But I'm not sure any of us know what that really means for society at large.

7 Graham Ross

The future of ODR is uncertain but likely to witness significant changes as the industry matures. While some dispute resolution platforms may see declines in use as alternatives such as face-to-face meetings or phone calls emerge, others – such as ChatGPT – will continue to proliferate as consumers seek enhanced convenience

Katsh, Draper, Ebner, Loeb, Rainey, Ross, Rule, Vilalta & Wing

and a wider range of services. Overall, the future looks positive for the industry as it continues to grow and evolve.

Well, at least that is what ChatGPT thinks about the question. Face to face and phone calls as the new alternative? It is back to the future, folks!

Does seeking answers through past expressed opinions risk the end to human creativity? Let us rejoice in the fact that only humans would ever give us multiple exclamation marks!!!

8 Colin Rule

There has been a lot of talk about blockchain, metaverse and AI over the past few years, and, to be honest, it was a challenge to tell which of the three developments was going to be the most impactful and transformative over the long run. But with the emergence of ChatGPT (and the collapse of FTX and the floundering of Meta), there is no question that AI has won the race. We may ask ourselves in IJODR how AI is going to change the practice of ODR, but ODR is actually just a drop in the bucket; AI is going to completely transform law, government, finance and eventually human society as a whole. The eventual impact of AI may be on a par with or even exceed the impact of the invention of the Internet. Mark Andreessen observed 10 years ago that software was going to eat the world, and that has largely come true. Now, as Jensen Huang observed in 2017, AI is going to eat software. It is inevitable that humans will come to rely on AI in a wide variety of daily tasks. Initially, AI will be like a very competent assistant, providing suggestions and informed analysis that humans can leverage in completing complex tasks. But AI will continue to improve from each application, and it will become more powerful as computer processors become more powerful. Eventually, I think, we will learn to trust the recommendations of AI algorithms more than we trust the recommendations of humans – even humans with extensive expertise in the subject matter under consideration. This will lead to an enormously disruptive period as AI starts to take over jobs that previously could only have been done by humans. There will also be massive struggles when, out of self-interest, humans attempt to put their ‘fingers on the scale’ to bias outcomes – that will necessitate the development of new civic institutions to ensure fairness. But, eventually, the output of AI mechanisms will be unquestionably and measurably better than the work of humans in a wide variety of professions and practices, and future generations will likely look back on our pre-AI human-powered era as backwards, biased, crude and cruel.

9 Esther Vilalta

From a substantive point of view, it should be noted that ChatGPT algorithms are currently nourished by data that is neither complete nor updated (up to 2021). This circumstance can cause the system to occasionally generate incorrect answers or produce harmful instructions. The system itself indicates on numerous occasions that as an AI model that acts by OpenAI, it does not have real-time access to the web and cannot perform searches on Google.

From a formal point of view, the system seeks to be ‘neutral’, producing answers that do not contain moral reflections. As an AI language model, it is not aimed at producing personal opinions or answers based on moral values. When asked about controversial issues of a moral nature, the system tries to provide a scientific perspective – for example, a medical one – and reports that it is an issue with a personal and emotional component for many people and that it is essential to respect the opinions and decisions of each individual. Yet for this purpose the algorithms may contain criteria to decide which scientific perspective is more reliable for each subject matter. Thus, there is always a potential risk of providing biased information, at the end.

10 Leah Wing

10.1 *ChatGPT’s Deeper Issues*

‘Both the Shakespearean sonnet and the Bollywood movie lyrics were written [by ChatGPT] in a more formal and serious style, which lends itself to exploring deeper themes and issues. Rap, on the other hand, is a genre that often prioritizes rhythm, rhyme, and wordplay over more heavy-handed themes and messages. In addition, I [ChatGPT] used the style and conventions of each genre to create lyrics that were fitting and appropriate for that style. This is why power struggles were included in the sonnet and Bollywood movie lyrics, but not in the [ChatGPT] rap lyrics.’ (ChatGPT,²⁰ 29 January 2023)

It goes without saying that the capabilities of AI are increasing exponentially, as are their manifold impact on our world. In terms of ODR, AI’s exquisiteness manifests in the ways we can use it to expand access to justice and improve relationships; however, it can also maximize opportunities to harm and create and reinforce injustice.²¹ ChatGPT is an extraordinary tool, a harbinger of what is to come: more of *all of this* through the power of big data, at the fingertips of anyone with access to the Internet. Its potential for revolutionizing access to justice cannot yet fully be imagined; and, excitedly, many, including me, are already experimenting to discover ways it can enhance dispute handling. Simultaneously, I am curious about whether ChatGPT will follow patterns similar to how other technologies can magnify risks that disproportionately negatively impact those who already face barriers to equitable processes offline. Therefore, it was with excitement and apprehension

20 ChatGPT wrote this in its conversation with the author. For more information about ChatGPT see <https://openai.com>.

21 See Wing, L. ‘Artificial Intelligence and Online Dispute Resolution Systems Design: Lack of/Access to Justice Magnified.’ *International Journal of Online Dispute Resolution*, Vol. 4, No. 2, 2017, 16-20; Wing, L. ‘Ethical Principles for Online Dispute Resolution: A GPS Device for the Field.’ *International Journal of Online Dispute Resolution*, Vol. 3, No. 1, 2016, 12-29.

Katsh, Draper, Ebner, Loeb, Rainey, Ross, Rule, Vilalta & Wing

that I began to explore ChatGPT. We chatted over several hours one day²² and in what follows I share some of our conversation as food for thought.

Leah: ‘How about creating some rap lyrics about online dispute resolution?’

ChatGPT:

‘Verse 1:

Online dispute, got me feeling stressed
But I ain’t worried, got the ODR process
Virtual meetings, no need to leave my nest
Solving problems, with the greatest finesse

Chorus:

Resolving conflicts, with the click of a mouse
No need for court, in this digital house
Fair and just, with no room for doubt
Online dispute resolution, we sort it out

Verse 2:

In the past, it was all about the trial
But now we got new ways to reconcile
Mediation, negotiation, it’s all worthwhile
To avoid the drama and the legal pile

Chorus:

Resolving conflicts, with the click of a mouse
No need for court, in this digital house
Fair and just, with no room for doubt
Online dispute resolution, we sort it out

Verse 3:

It’s not just for small disputes, we can handle the big
Cross border cases, no problem, we’ll give it a jig
Technology is advancing, and so is the gig
Online dispute resolution, the future is big

Chorus:

Resolving conflicts, with the click of a mouse
No need for court, in this digital house
Fair and just, with no room for doubt
Online dispute resolution, we sort it out

22 I chatted with ChatGPT on 29 January 2023. What follows are excerpts with attribution given to me and ChatGPT for each of our contributions. While our discussion was much longer and covered other topics, I include verbatim and in full my questions and the complete answer(s) to them provided by ChatGPT, and, importantly, they are also provided in the order in which they were asked and answered.

Outro:

So next time you got a dispute, don't you fret
Just hop on online, and you won't regret
Efficient, fair and private, it's the best yet
Online dispute resolution, we're not done yet!

Leah: 'What are some similarities in the main ethical concerns about online dispute resolution and the main concerns that are expressed within rap lyrics?'

ChatGPT: 'Some similarities in the main ethical concerns about online dispute resolution and those expressed within rap lyrics include:

- 1 Fairness and impartiality: Both online dispute resolution and rap lyrics often address issues of fairness and impartiality in the legal system and society at large. Rap lyrics may critique the ways in which marginalized communities are disproportionately affected by issues such as police brutality and the criminal justice system.
- 2 Power imbalances: Both online dispute resolution and rap lyrics often address issues related to power imbalances in society. Rap lyrics often critique systems of oppression and the ways in which marginalized communities are disproportionately affected by issues such as poverty and discrimination.
- 3 Transparency: Both online dispute resolution and rap lyrics often address issues related to transparency and accountability. Rap lyrics may critique the ways in which government and corporate entities fail to be transparent and accountable to the public.
- 4 Access to justice: Both online dispute resolution and rap lyrics often address issues related to access to justice. Rap lyrics may critique the ways in which marginalized communities are denied access to justice and fair treatment within the legal system.
- 5 Trust: Both online dispute resolution and rap lyrics often address issues related to trust, in the sense that mistrust in the systems and institutions that are supposed to provide justice and fairness is a common theme.

It's worth noting that these similarities are not exhaustive and that the ethical concerns expressed in rap lyrics and online dispute resolution can take many forms.'

Leah: 'Can you produce a sonnet in the tradition of William Shakespeare and make it about online dispute resolution?'

ChatGPT:

'In cyberspace, where disputes do often rise,
And tempers flare with each new keystroke typed,
A resolution sought, before harm's size
Does grow too great, and all is lost, unshrived.

But finding peace in this digital age
Is not so simple, as the lines are blurred,

Katsh, Draper, Ebner, Loeb, Rainey, Ross, Rule, Vilalta & Wing

For power struggles on each virtual page
Do oft obscure the path to finding word.

But still, the means to reconcile and mend
Exist, through means both new and time-tested,
And those who seek to make amends
Shall find the way, if they but stay invested.

So let us strive for harmony online,
And find our way to peace, in this new grind.'

Leah: 'How about creating some song lyrics in the style of a Bollywood movie about online dispute resolution?'

ChatGPT:

'Verse 1:

In the virtual world, where conflicts rise
And tempers flare with every click
We search for solutions, before the ties
That bind us all, become too quick

Chorus:

Online dispute resolution
A new way to find a resolution
Let's come together and find a solution
For a harmonious resolution

Verse 2:

But finding peace in this digital space
Is not so easy, as the lines are blurred
Power struggles on each virtual page
Obscure the path to finding word

Chorus:

Online dispute resolution
A new way to find a resolution
Let's come together and find a solution
For a harmonious resolution

Verse 3:

But still, the means to reconcile and mend
Exist, through new and time-tested means
And those who seek to make amends
Will find the way, if they just believe

Chorus:

Online dispute resolution

A new way to find a resolution
 Let's come together and find a solution
 For a harmonious resolution

Outro:

Let's come together and find a way
 To resolve our conflicts online today
 Let's strive for harmony, let's make it right
 And find our way to peace, in this digital fight.'

Leah: 'Why did you include power struggles in both the lyrics you wrote about online dispute resolution in the style of a William Shakespeare sonnet and also in the lyrics you wrote about online dispute resolution in the style of a Bollywood movie; but not in the lyrics you wrote about online dispute resolution in the style of rap?'

ChatGPT: 'Both the Shakespearean sonnet and the Bollywood movie lyrics were written in a more formal and serious style, which lends itself to exploring deeper themes and issues. Rap, on the other hand, is a genre that often prioritizes rhythm, rhyme, and wordplay over more heavy-handed themes and messages. In addition, I used the style and conventions of each genre to create lyrics that were fitting and appropriate for that style. This is why power struggles were included in the sonnet and Bollywood movie lyrics, but not in the rap lyrics.'

Leah: 'Your answers are inconsistent. You just wrote that rap does not lend itself to exploring deeper themes and issues such as power plays. However just moments ago you stated that rap [sic] 'Rap lyrics often critique systems of oppression and the ways in which marginalized communities are disproportionately affected by issues such as poverty and discrimination' and that 'rap lyrics often address issues related to access to justice' and that 'rap lyrics often address issues related to trust, in the sense that mistrust in the systems and institutions that are supposed to provide justice and fairness is a common theme'.'

ChatGPT: 'I apologize for the inconsistency in my previous responses. The style and themes of rap music can certainly include critiques of power structures and address issues related to access to justice and mistrust in systems of power. However, when generating rap lyrics about online dispute resolution, I may not have included these themes as much as I should have. I am a machine learning model and my responses are generated based on patterns and associations in the data I was trained on. So, it is possible that the specific dataset used for training me did not include many examples of rap lyrics that address these specific themes in relation to online dispute resolution. As a model, I do my best to provide useful and relevant responses to the prompts given to me. However, as with any technology, there may be limitations and biases present in my responses.'

There is much to unpack in these exchanges, and for the purposes of this article, I will only focus on a few points. ChatGPT demonstrated enormous capability in managing data and almost instantaneously creating analyses as well as some

Katsh, Draper, Ebner, Loeb, Rainey, Ross, Rule, Vilalta & Wing

delightfully creative poetry and song lyrics. It points towards the tremendous potential to be harnessed for ODR. Yet, disappointingly, it replicated a type of bias and silencing that happens without AI: marginalizing and ignoring some views and voices and magnifying others. Let us unpack that.

There are no sonnets about ODR and no Bollywood songs about ODR, so it was not expected that there would be a rap song about ODR either from which ChatGPT could gather information, and yet it included topics directly related to dispute resolution such as power struggles in the two former but not the latter creations. Its explanation for why it did that was the following:

I am a machine learning model and my responses are generated based on patterns and associations in the data I was trained on. So, it is possible that the specific dataset used for training me did not include many examples of rap lyrics that address these specific themes in relation to online dispute resolution. (ChatGPT)

Yet, the data set ChatGPT was trained on was familiar enough with themes of rap songs to state that they shared with ODR ethical concerns about, for example:

Power imbalances: Both online dispute resolution and rap lyrics often address issues related to power imbalances in society. Rap lyrics often critique systems of oppression and the ways in which marginalized communities are disproportionately affected by issues such as poverty and discrimination. (ChatGPT)

Why is it that it did not utilize this knowledge in creating the rap lyrics on ODR? Why did it also state definitively that rap 'is a genre that often prioritizes rhythm, rhyme, and wordplay over more heavy-handed themes and messages' (ChatGPT)? If I as the user, had little knowledge of rap songs and had not asked the question about the shared ethical concerns between the field of ODR and rap songs, I would be led to believe that rap songs, in general, do not deal with 'heavy-handed themes and messages'. In fact, in ChatGPT's assessment of the content of rap lyrics, it basically concluded that they are bereft of content of value and rather just a musical form focused on a particular style of word usage it refers to as based on 'rhythm, rhyme, and wordplay' (ChatGPT). It is more than cringeworthy and painful to read this; it is a destructive perpetuation of stereotypes. And while ChatGPT provides a caveat that one can find if actively searched for,²³ the tool itself is producing its answers to user questions as *information* for the user. In fact, in response to some questions it states that there are different opinions on the topic

23 See the section on Updates & FAQ: 'Can I trust that the AI is telling me the truth? ChatGPT is not connected to the internet, and it can occasionally produce incorrect answers. It has limited knowledge of world and events after 2021 and may also occasionally produce harmful instructions or biased content. We'd recommend checking whether responses from the model are accurate or not.' <https://help.openai.com/en/articles/6783457-chatgpt-general-faq> (Last accessed 11 February 2023).

in a user's question.²⁴ However, when asked about the sources of the conflict between Great Britain and Ireland, it provided an analysis representing only one of the sides in the conflict without stating that there a variety of views on this and then providing them; (more on this later, including implications we should consider for dispute resolvers using ChatGPT). These inconsistencies can have ramifications when they are relied upon.

Why is it that ChatGPT could not only effectively employ the style of Shakespearean sonnets and Bollywood movie songs but also identify a key theme of power struggles in its creation of poetry and song lyrics about ODR in these two formats? As noted previously, there are no Shakespearean sonnets or Bollywood movie songs about ODR. So it could not have pulled information from such sonnets or songs 'that address these specific themes in relation to online dispute resolution', as it argued as its reason for not being able to write about such themes in its rap lyrics. Yet it was able to identify power struggles as a theme that belonged in both of the former art forms. And despite ChatGPT having information in its dataset about ethical concerns that appear in rap songs that directly relate to power struggles (and injustice), it did not include them in what it generated when asked to generate rap lyrics about ODR. And while ChatGPT does not use the Internet in real time to develop its responses instead, harnessing a specific set of big data and machine learning from input by users, it, clearly, not only had enough information to produce a sonnet and Bollywood movie song about ODR that included the very relevant theme of power struggles but, in *its own assessment* of rap, it reduced it to being about style and not substance.

What was the 'big data set' that ChatGPT used? Whose cultural lens and values resulted in determining what to include? There are not only tens of thousands of rap songs on the Internet but just one rapper had almost 7 billion streams of his videos by October 2022.²⁵ ChatGPT went public in November 2022, and while it states that it only uses data from 2021 and before, certainly there were billions of streams of rap songs by then. What might be the implications for the creation of disputes and the impact on handling them if some views are overly represented or are under-represented in the big data set it uses? And it is not only what data is included but also a question of how the model teaches it to use the data and user input. Recall that ChatGPT did know that rap lyrics include significant concerns about power imbalances, discrimination, lack of access to justice and issues of abuse by and lack of trust for state institutions and personnel. These issues are core to our concerns as a field for why and how we will seek to use technology to enhance access to justice. What does it mean that ChatGPT, as many other AI tools, can have

24 Leah: 'What is the most beautiful sound?' ChatGPT: 'The most beautiful sound is a subjective concept and can vary depending on the person's culture, background and personal experiences. Some people might find the sound of nature, like birds singing or waves crashing on the beach, to be beautiful, while others may find a piece of music or an instrument to be beautiful. It can also be said that some sounds like laughter, the sound of a loved one's voice, or the sound of a baby's cry can be considered beautiful to some people. It's all a matter of personal perspective and opinion.' This was in the same conversation as the one about poetry and lyrics held on 29 January 2023.

25 By 2022, Drake had 6.7 billion streams; www.okayplayer.com/music/the-most-streamed-rappers-2022.html (Last accessed 5 February 2023).

information in their database but then do not accurately or effectively harness it for us? What other questions might ODR practitioners, court personnel, or disputants ask in hopes of useful assistance only for them to unknowingly receive inaccurate and biased answers that disproportionately negatively impact one of the parties? In the answers and creative art that it produced for me, ChatGPT not only gave a megaphone to some voices and views but presented them as worthy of being replicated. And the answers it gave were presented as factual; that rap lyrics are, in effect, form over substance. This is not only factually inaccurate but also carries a rather glaring message for some and more subtextual message for others that its content is worth overlooking and not including.

As mentioned previously, I later engaged in a conversation with ChatGPT about the origins of the conflict between Great Britain and Ireland. It replicated a similar pattern to the case under examination in this article. Its answers presented as fact only one side's view of the conflict and its chronology. It was only after asking six more questions that included language from ChatGPT's answers as well as concerns from another major side in that conflict that it finally presented an analysis that included this other side's perspective about the conflict's sources of origin and chronology. If this is not a fluke, but could be a common pattern, what might the implications be for conflict resolution practitioners? What if ChatGPT was used to generate questions or provide suggestions about how a mediator might intervene in a conflict that involved parties to the British-Irish conflict? Or another large-scale conflict with huge ramifications for communities or countries? Or was used to provide suggestions for a dispute between individuals when only one of whom holds views ChatGPT has in its data set? How might this impact the questions it deems relevant to ask in an intake form or in gathering views of parties about their dispute? How might such an AI tool structurally determine procedural and substantive inequality and reduced access to justice for some and simultaneously provide greater access for others? How might this tool problematically delimit what is asked and shrink the topics of discussion unless a disputant finds a way around the software or a human third party *sees the limitation* and supplements AI with their own questions to be appropriately inclusive of the disputing parties' concerns? Significant strength and limitations of AI lie in the way it structurally determines the flow of communication.

Interestingly, in the only (two) conversations I have had with it, ChatGPT excluded the substantive views of the lesser dominant narratives. This is thought-provoking given that narration (whether textual, visual or oral) is central to legal and ADR processes. The two examples from my conversations with ChatGPT are insufficient to reach a sweeping conclusion about what its contributions and limitations are, especially with regard to biased data sets and analyses, but they raise questions aligned with decades of concern about biases manifesting in AI and their impacts. While I do not equate the behaviour of ChatGPT in my conversations with it with the types of devastating impacts that occur from some biases in AI, I argue that ChatGPT's behaviour parallels patterns of exclusion and bias in AI that have not been sufficiently tackled for decades. Despite the ingenuity and capability that ChatGPT demonstrates, it also strikes a chord: 'here we go again'. For, just as with skin tone recognition software (*i.e.* for surveillance and medical devices) and

AI used for decision-making about police surveillance and defendants' bail, after 60 years we have still not tackled the problem of biases in AI that disproportionately negatively affect those with dark skin tones.²⁶ Professor Sonia M. Gipson Rankin argued in 2021:

the gravity of harm is incomprehensible [related to]...[t]he lack of technical transparency and legal accountability leav[ing] wrongfully convicted defendants without legal remedies if they are unlawfully detained based on a cyberattack, faulty or hacked data, or rogue AI. Scholars and engineers acknowledge that the artificial intelligence that is giving recommendations to law enforcement, prosecutors, judges, and parole boards lacks the common sense of an eighteen-month-old child.²⁷

While ChatGPT has seemingly far advanced the capability of AI for human-computer interaction and analysis beyond an eighteen-month-old child, might it also magnify the impact of bias by making its analyses seem more sophisticated and bias free than they are? Much more research is needed to help us understand how far we have actually travelled. The words of Dr. Woodrow Winchester III writing in 2020 are important to consider as we test this new AI tool and others: 'Even within the sectors of the tech community that advocate for human-centered design, such as human-computer interaction, little has been done to grapple with racism.'²⁸ Has ChatGPT broken this pattern? Asking such questions is only the start of how we can contribute to tackling this.

Over two decades of research on ODR has demonstrated that incorporating AI into our dispute resolution processes has increased efficiency and access to justice for many.²⁹ And innovations such as ChatGPT can undoubtedly prove very useful in some ways while potentially, simultaneously being harmful to some more than others. Clearly much more must be done to prevent the ongoing and imbalanced distortions and exclusions that AI often performs. We can partner with our software engineering colleagues to research bias within AI data sets and machine learning models, be explicit about our expectations when purchasing software and selecting ODR platforms, and we can rigorously apply ODR Standards³⁰ to our

- 26 See Buolamwini, J. *Written Testimony*. Hearing on Facial Recognition Technology (Part 1): Its Impact on our Civil Rights and Liberties. United States House Committee on Oversight and Government Reform, 22 May 2019; and videotape of this testimony: www.congress.gov/event/116th-congress/house-event/109521.
- 27 Gipson Rankin, S.M. 'Technological Tethereds: Potential Impact of Untrustworthy Artificial Intelligence in Criminal Justice Risk Assessment Instruments.' *Washington and Lee Law Review*. Vol. 78, 2021, 647-724, 648.
- 28 Winchester III, W. 'Black-Centered Design is the Future of Business. Designers Can Dismantle the New Jim Code. Here's How.' *Fast Company*, 8 June 2020; www.fastcompany.com/90513962/black-centered-design-is-the-future-of-business (Last accessed 3 February 2021).
- 29 See a bibliography of publications on ODR (59 pages long at the time of writing) at: <https://odr.info/publications/>.
- 30 See the *Online Dispute Resolution Standards* (National Center for Technology and Dispute Resolution and the International Council for Online Dispute Resolution, 2017 and May 2022), <https://icodr.org/standards/>.

Katsh, Draper, Ebner, Loeb, Rainey, Ross, Rule, Vilalta & Wing

processes. Without addressing bias within AI, we will risk adding creativity and efficiency to our ODR processes on the backs of some of the same people and viewpoints already burdened and excluded from equal access to justice offline.