

# The Pull of Unbiased AI Mediators

Chris Draper\*

## Abstract

*There is significant concern in the access to justice community that expanding current court-based online dispute resolution (ODR) efforts will further exacerbate the systemic inequities present in the American justice system. This well-founded fear stems from the fact that current ODR tools typically calibrate artificial intelligence (AI) algorithms with past outcomes so that any future cases are consistently analysed and filtered in a manner that produces similar results. As courts consider ODR tools for more complicated cases that often require mediation, there is significant disagreement on whether it is possible to create an AI mediator and how that could be achieved. This article argues that an effective AI mediator could be created if its design focuses not on the outcomes achieved by the mediation but on the manner of the communication prompts used by the AI mediator.*

**Keywords:** automation, artificial intelligence, algorithm development, mediation, pull style communication.

## 1 Introduction

The response to automation in traditionally white-collar industries ranges from excitement to fear.<sup>1</sup> Online Dispute Resolution (ODR) is one of these primarily white-collar applications where the onslaught of automation is causing a wide range of visceral reactions. To some, implementing ODR solutions that “turn every cellphone into a point of access to justice” is the only way we can meaningfully serve Americans who cannot afford a lawyer when they need one.<sup>2</sup> For others, ODR tools are accelerating efforts to remove all remaining shreds of humanity from our inequitable society.

\* Chris Draper, PhD, PE, is the Managing Director of Trokt, responsible for guiding the development, adoption and growth of the Trokt Online Dispute Resolution platform. Dr. Draper is a trained engineer with a focus on human-technology interface risks, a certified mediator with a focus on special needs education conflicts, and an expert on the evaluation of highly complex systems that assist in the human management of legally sensitive data. Dr. Draper received his Bachelor of Science from the University of California at Berkeley and his Doctor of Philosophy from the University of Glasgow.

1 See, ‘Automation and Anxiety’, *The Economist*, 23 June 2016, [www.economist.com/special-report/2016/06/23/automation-and-anxiety](http://www.economist.com/special-report/2016/06/23/automation-and-anxiety) (last accessed 7 July 2019).

2 A. Morris, ‘Could 80 Percent of Cases Be Resolved through Online Dispute Resolution?’ *Legal Rebels Podcast*, 17 October 2018, [www.abajournal.com/legalrebels/article/rebels\\_podcast\\_episode\\_033](http://www.abajournal.com/legalrebels/article/rebels_podcast_episode_033) (last accessed 7 July 2019).

Alternative dispute resolution (ADR) covers a wide range of dispute resolution techniques designed to streamline justice by avoiding traditional courts. ODR takes efficiency a step farther by streamlining traditionally human-dependent processes. For example, traditional ADR expects mediation to bring all parties to one location where information can be physically shuttled between different physical locations. ODR mediation can be carried out asynchronously with electronic communication between parties from all over the world securely facilitated through a platform like Trokt. While there is limited controversy when ODR automates the physical message delivery happening in traditional ADR, artificial intelligence (AI) is being looked to for automating procedural and facilitative processes that could significantly impact ODR outcomes.

The most controversial conceptual application of automation is an *AI mediator*. There is a body of evidence that indicates AI is fundamentally unable to be truly creative<sup>3</sup> or to make decisions that will be acceptable to humans without continually optimizing human input.<sup>4</sup> These fundamental issues with AI are often cited as reasons why an AI mediator is not viable. Yet these arguments implicitly assume that mediation requires creative skill.

## 2 What If It Does Not?

There are disputes where the presence of a specific individual can fundamentally alter the outcome of the dispute or any agreement that could be reached. Yet there does appear to be a science to mediation that could be repeatably designed. If true, then a thoughtful development and application of ODR technologies, contextualized by the actual risks and biases observed in unautomated ADR workflows, could provide a future that is squarely in the middle of these two extreme visions – even when it comes to an AI mediator.

## 3 What Is ODR?

ODR is fundamentally an ADR process that is at least partially completed online. Dispute resolution processes include fact-finding, negotiation or settlement during mediation, arbitration or med-arb using video, voice or written collaboration software, to name just a few. At its most basic, ODR includes asynchronous contract negotiation over email, or real-time collaboration using text-based messaging programmes. At its more refined, ODR includes case management software that is integrated into court processes, or collaboration spaces that allow brainstorming to feed into document drafting until an agreed contract is signed.

3 S. Belsky, 'How to Thwart the Robots: Unabashed Creativity', *Fast Company*, 23 January 2019, [www.fastcompany.com/90294821/how-to-thwart-the-robots-unabashed-creativity](http://www.fastcompany.com/90294821/how-to-thwart-the-robots-unabashed-creativity) (last accessed 7 July 2019)

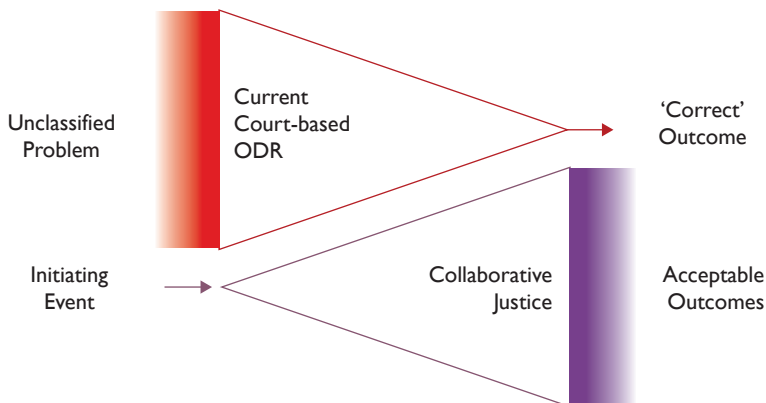
4 N. Scheiber, 'High-Skilled White-Collar Work? Machines Can Do That, Too', *New York Times*, 7 July 2018, [www.nytimes.com/2018/07/07/business/economy/algorithm-fashion-jobs.html](http://www.nytimes.com/2018/07/07/business/economy/algorithm-fashion-jobs.html) (last accessed 7 July 2019).

Chris Draper

Regardless of the ODR platform's sophistication, it will typically take one of two operational forms: filtering or facilitating.

Most discussion around ODR often focuses on court-based systems integrating tools from companies like Tyler Technologies or Matterhorn to solve simple, low-value disputes.<sup>5</sup> If an individual receives a parking ticket, these tools allow someone to log in, complete the necessary documents and pay all relevant fines online. On the more complex end, these tools may allow for the solving of small claims by facilitating real-time or asynchronous text-based communication between the parties, allow the claimant to accept a lesser value in exchange for a quick resolution, and all agreements and payments will be completed online. In all these types of cases, the number of potential outcomes is limited. Whether it be a traffic ticket or a small-claim dispute, the claimant will get all its money, some of its money, none of its money or go to court. These 'filtering' ODR systems are designed to filter the parties into one of these outcomes as quickly as possible. Filtering ODR systems are represented by the top part of Figure 1, where an unclassified problem is filtered down to the 'correct' outcome.

**Figure 1** Filtering versus Facilitating ODR Systems



The less visible branch of ODR is 'facilitating' tools that fall into the realm of collaborative justice. These tools are designed to ensure that any outcome that is acceptable to all involved parties can be simply, safely and securely codified. The most widely recognized ODR tool in this space is Trokt, which is designed to remove the human errors that occur between the time a dispute arises until a settlement agreement is filed. Yet Trokt is not the market leader. The US district courts terminated 98,786 tort trial cases between 2002 and 2003, with only

5 J.J. Prescott, 'Improving Access to Justice in State Courts with Platform Technology', *Vanderbilt Law Review*, Vol. 70, No. 6, 1993, <https://s3.amazonaws.com/vu-wp0/wp-content/uploads/sites/89/2017/11/28175541/Improving-Access-to-Justice-in-State-Courts-with-Platform-Technology.pdf> (last accessed 7 July 2019).

1,647 of these cases resulting in either a jury or bench trial.<sup>6</sup> The remaining 98.3% of tort cases that did not go to trial were resolved with some mixture of the most widely used facilitating ODR tools: tracked-changed Word documents going back and forth via email that are discussed over the phone or in a videoconference. Despite the widespread understanding that these types of tools are not secure and are inappropriate for relaying sensitive digital data,<sup>7</sup> tens of thousands of complex settlements are negotiated online every year using these rudimentary ODR tools.

The case management and negotiation tools developed by Tyler and Matterhorn have recently defined what many consider court-based ODR. Yet the fundamental definition of ODR means our justice system has adopted and accepted the use of ODR tools for decades. The emerging concerns about ODR regard how these tools are becoming automated.

#### 4 Acceleration through Automation

Current court-based ODR tools such as those provided by Tyler and Matterhorn that rely on ‘filtering’ a dispute into its appropriate resolution bin achieve efficiency by increasing the speed at which a case is filtered. Increasing filtering speed means users must:

- agree to use the system more quickly,
- acknowledge they understand the process more quickly,
- be appropriately filtered into an agreeable bin more quickly,
- agree to the terms of the resolution accessible in the bin where they are filtered more quickly, and
- complete the actions required by the agreement more quickly.

These steps can be achieved more quickly by:

- Simplifying descriptive or instructional language,
- Classifying the dispute with AI, and
- Reducing the effort of agreeing to and paying for the resolution.

These same filtering concepts are routinely applied to the commonly used ‘facilitating’ tools of Word, email and text-based messaging during eDiscovery. Most common eDiscovery tools are using Natural Language Processing (NLP) AI methods to find words, phrases or concepts within documents being searched and select what documents or data points the eDiscovery tool ‘believes’ are appropriate for human review as part of the discovery effort. These same types of AI tools are already appearing in email and word processing tools to help users select words that adjust how the reader will interpret things like repetitive content or intended tone.

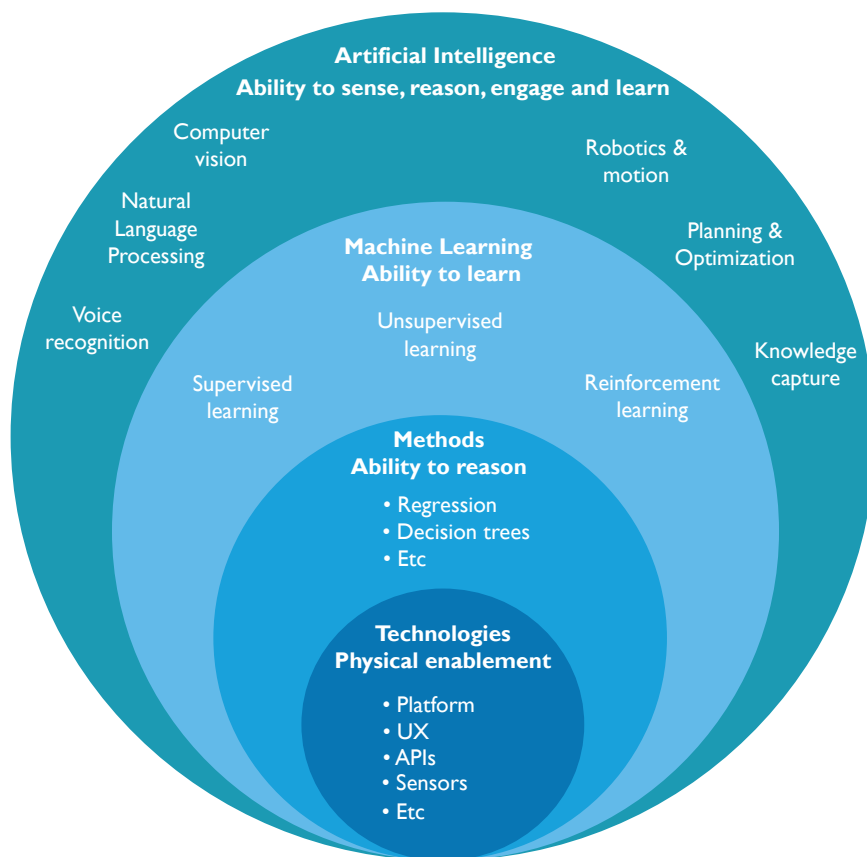
6 See, ‘Tort, Contract and Real Property Trials’, *Bureau of Justice Statistics*, [www.bjs.gov/index.cfm?ty=tp&tid=451](http://www.bjs.gov/index.cfm?ty=tp&tid=451) (last accessed 7 July 2019).

7 C.H. Draper & A.H. Raymond, ‘Building a Risk Model for Data Incidents: A Guide to Assist Businesses in Making Ethical Data Decisions’, *Business Horizons*, 2019, ISSN: 0007-6813, <https://doi.org/10.1016/j.bushor.2019.04.005> (last accessed 7 July 2019).

Chris Draper

Not only is the use of these types of AI techniques for both ‘filtering’ and ‘facilitating’ ODR tools common, but it enables users to rapidly find consistency within large or often unstructured datasets. This consistency is found by using Machine Learning (ML) algorithms that compare large amounts of Big Data (BD) against Business Intelligence (BI) metrics that set the priorities for comparison. If these priorities for comparison are met, whether the AI process is relevant to NLP, Vision, Autonomous Vehicles, Robotics or other common applications, the system takes the action that it is instructed to take when faced with this consistency.<sup>8</sup>

Figure 2 Artificial Intelligence, Dependent Tools and Technologies<sup>9</sup>



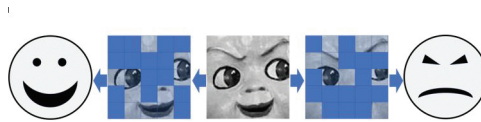
8 S. van Duin & N. Bakhshi, ‘Artificial Intelligence Defined’, *Deloitte*, March 2017, [www2.deloitte.com/se/sv/pages/technology/articles/part1-artificial-intelligence-defined.html](http://www2.deloitte.com/se/sv/pages/technology/articles/part1-artificial-intelligence-defined.html) (last accessed 7 July 2019).

9 S. van Duin & N. Bakhshi, ‘Artificial Intelligence Defined’, *Deloitte*, March 2017, [www2.deloitte.com/se/sv/pages/technology/articles/part1-artificial-intelligence-defined.html](http://www2.deloitte.com/se/sv/pages/technology/articles/part1-artificial-intelligence-defined.html) (last accessed 7 July 2019).

For example, an autonomous vehicle application of AI is often seeking unobstructed road and guiding the vehicle to maximize its allowed speed along that unobstructed road. Alternatively, a Vision application of AI may be operating at a retail store to identify how many items are on a shelf before and after a customer enters a store and to alert security if the number of items remaining is not consistent with the number of items bought. Or a Robotics application of AI may be testing the ripeness of grapes by tracking the response to compression forces and picking those that conform to customer interpretations of preferred taste. All these applications depend on ML to process the data the software is gathering against the data with which the ML routines are calibrated, with AI considered to have the ability to make the 'correct' decision with this processed data.

The fundamental limitations of AI are no different from those we experience as humans. First, the quality of any AI application is directly dependent on its ability to access large amounts of relevant data. When data is limited, we must make assumptions about what we cannot see. Second, the quality of any AI decision is directly dependent on its ability to accurately equate what it sees with reality. When environments change, there is an inertia to our expectation of human reactions. This is easy to see in the facial recognition challenges like those presented in Figure 3.

**Figure 3** Example of the Interpretation Challenges Given Limited Data



In Figure 3, we see that different limited data views result in wildly different interpretations of the underlying reality. Yet even with the full picture, the underlying emotional reality may remain unclear. This lack of clarity and the method of reducing it is not unique to facial recognition, where the interpretation of meaning in facial expressions is continually studied using large human surveys.<sup>10</sup> These large surveys often calibrate a 'Weak' or 'Narrow' AI system that will use this data as the basis for making its decisions. These Weak or Narrow AI systems differ from 'Strong' AI systems that independently refine their analysis of the data, the most important difference being that *Strong AI systems do not currently exist*.

It is extremely important to understand two things when interpreting AI systems:

- 1 AI does not uncover truth; it more rapidly refines what it sees, and
- 2 The weaknesses of AI systems are equivalent to those in an expert witness.

10 S. Fecht, 'How to Smile without Looking Like a Creep, According to Scientists', *Popular Science*, 28 June 2017, [www.popsi.com/how-to-smile](http://www.popsi.com/how-to-smile) (last accessed 7 July 2019).

Chris Draper

These truths about AI systems are important to understand when interpreting, assessing and mitigating the risks around their use in access to justice applications.

## 5 What Aspects of ODR Worry the Access to Justice Community?

In the dispute resolution and access to justice communities, court-based ODR processes raise a number of concerns and fears. These concerns and fears about court-based ODR technology in general and AI-assisted technology more generally included eight major concerns that were expressed at the National Legal Aid & Defender Association (NLADA) Tech Section meeting held during the 2019 Equal Justice Conference in Louisville, Kentucky.<sup>11</sup> These eight issues that participants at the Tech Section believe must be accounted for in court-based ODR tools can be summarized and grouped as follows:

### Group A: Uniform Intent

1. Stakeholder participation. Automating significant portions of the court system using ODR tools could produce wide-ranging and potentially unintended consequences, including the dominance of systemic biases. For this reason, it is important that the broadest collection of impacted stakeholders be involved in the planning and implementation of any changes.

### Group B: Safely Accessible

2. Equitable Accessibility. Internet access is neither of consistent quality nor equitably distributed throughout the United States. For this reason, the bandwidth and autosaving qualities of any cloud-based system should be understood and appropriate for the communities intended to access any ODR tool.

3. Physical Security. The ability to provide access to the courts from nearly any location using ODR tools means an individual could be completing a legally binding action in physical proximity to someone causing him or her duress. For this reason, scenarios that involve actions taken under duress that may not be possible without the presence of an ODR tool should be accounted for when determining the enforceability of any ODR-enabled agreement.

### Group C: Meaningful Clarity

4. Accurate Translations. The American judicial system must be accessible to individuals who are unable to fully understand the English language. For this reason, the implementation of any ODR system should not deny equitable access to individuals based on the languages they can understand.

5. Oversimplification of plain language. Many concepts, procedures or choices in the American justice system are complex enough that their implications cannot be condensed into simple summaries or binary choices. For this reason, informational or expert systems that accompany ODR systems, or the choices users are forced to make as part of an ODR system, should be careful to avoid oversimplifi-

11 C.H. Draper, 'Personal Notes from the NLADA Tech Section Meeting', *2019 Equal Justice Conference* in Louisville, KY, 9 May 2019.

cations that allow individuals to be adversely impacted by negative consequences that they did not expect or understand.

6. **Meaningful Allocation.** There is now a societal comfort with accepting terms and conditions couched in long, intellectually impenetrable language when using software platforms that results in the average user being completely unaware of their rights or responsibilities with respect to their use of those platforms. For this reason, any ODR tool should ensure that any meaningful, legal obligations associated with a process that is agreed to using the ODR tool are not bundled or buried in a manner that produces uninformed acceptance.

#### Group D: Information Security

7. **Operational Privacy.** Societal comfort with facilitative ODR tools makes most practitioners unaware that the operational risks associated with an approved user mismanaging data are conservatively estimated as eight times more likely to cause a release of confidential data than any malicious system intrusion.<sup>12</sup> For this reason, any court-based ODR tool should assess the operational risks associated with using the platform against the value of any data that a user could accidentally release.

8. **Anonymous Calibration.** The source and quality of the data required to calibrate an AI tool will fundamentally impact both the decisions made by the tool and the ability to rebuild private qualities defining the user whose data was anonymized for inclusion in the calibrating dataset. For this reason, the data selection, storage and calibration processes associated with an ODR tool should account for worst-case, real world impacts when assessing appropriateness.

Examining these concerns may bring many to one rather surprising conclusion: very few are related specifically to the ODR technology itself. For example, operational changes in court systems that are perceived as minor, like altering the reporting structure of administrative staff, can often result in significant impacts because not all stakeholders are included in the planning process. In the case of equitable access, the location, hours of operation and available transportation to a traditional courthouse may not provide an appropriate level of access or safety for many who need to participate in programmes located at the courthouse. In the case of meaningful clarity, ineffective or oversimplified written communication may even produce more significantly negative impacts when delivered in paper form because searchability and comparability are more limited relative to electronic documents that can be immediately linked to a wider array of more extensive resources. Or in the case of information security, the release risks that are often feared with digital data have no meaningful ability to even be tracked when a well-intentioned individual inappropriately shares paper-based materials where the transmission of digital data does have a reasonable ability to be meaningfully controlled.

12 C.H. Draper & A.H. Raymond, 'Building a Risk Model for Data Incidents: A Guide to Assist Businesses in Making Ethical Data Decisions', *Business Horizons*, 2019, ISSN: 0007-6813, <https://doi.org/10.1016/j.bushor.2019.04.005> (last accessed 7 July 2019).



Chris Draper

For those items that are related to the technologies behind the ODR tools, “reliance on algorithms and data, present new challenges to fairness and open the door to new sources of danger for disputants and the judicial system”.<sup>13</sup> For court-based ODR systems that rely on filtering towards predefined options, the technological risks are both clear and nearly unavoidable: the biases and inequities present in the data of past actions that are used to calibrate the future decisions of AI-based ODR tools will more perfectly replicate past biases and inequities more quickly.

ODR processes that rely upon facilitation, however, can either be equivalently harmed or uniquely enabled by AI. When disputes are resolved by allowing the parties to voluntarily arrive at a legally acceptable outcome that is agreeable to all parties, even if the final terms of the agreement were not initially contemplated as a potential outcome, there is greater access to more efficient, enforceable justice. ODR systems could achieve these types of outcomes when they are built to mimic effective mediation, with AI enabling an idealized mediator.

## 6 What Is Mediation?

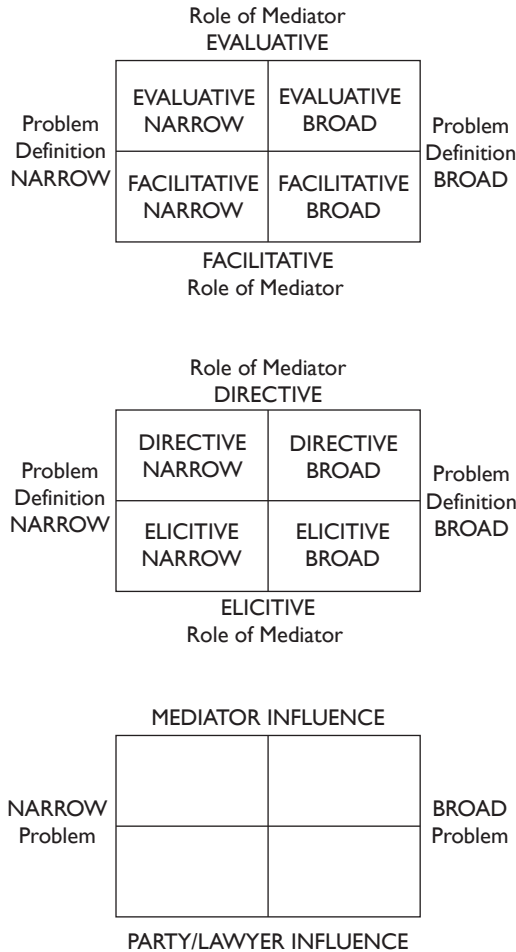
Mediation can be generally defined as a process where an impartial facilitator assists disputing parties to develop a mutually agreeable resolution. It is broadly accepted that the disputing parties are the only ones who can start or end a mediation, and a facilitating mediator must not have any interest in either side reaching any specific outcome. How the mediator assists the parties is a bit murkier.

Every mediation is a multiparty negotiation where the mediator is actively involved in the crafting of each party’s proposals. The nature of this involvement will typically vary over the course of a mediation, with many of the recent methods for classifying a mediator’s role influenced by the work by Leonard Riskin. Figure 4 shows three versions of Riskin Grids discussed in his evaluation of their intended use and effectiveness, with each defining the mediator’s role as a combination of its facilitative scope and influence.<sup>14</sup> This definition of mediator actions as represented by Riskin Grids has often resulted in conversations about automation that focus on the procedural or evaluative aspects of a mediation. In this regard, AI has already proved to be more effective than a human mediator.

13 O. Rabinovch-Einy & E. Katsh, ‘The New Courts’, *American University Law Review*, Vol. 67, p. 165, [www.aulawreview.org/au\\_law\\_review/wp-content/uploads/2017/12/03-RabinovchEinyKatsh.to\\_Printer.pdf](http://www.aulawreview.org/au_law_review/wp-content/uploads/2017/12/03-RabinovchEinyKatsh.to_Printer.pdf) (last accessed 7 July 2019).

14 L. Riskin, ‘Decisionmaking in Mediation: The New Old Grid and the New New Grid System’, *Notre Dame Law Review*, Vol. 79, No. 1, Article 1, 1 December 2003, <https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=1416&context=ndlr> (last accessed 7 July 2019).

**Figure 4** *Various Versions of Riskin Grids*



For example, if the parties wish to constrain the mediator’s focus to a narrow aspect of a complex topic, an AI system could perfectly exclude all data deemed out of scope by the parties without any residual bias. Alternatively, if the parties are seeking an evaluation of how a particular aspect of the negotiation would likely be resolved on the basis of its similarity to all other known disputes, AI can make a more accurate assessment more efficiently than any human. These uses of AI are already proven tools for filtering a dispute and are exactly the kind of AI applications that run the risk of more quickly replicating past biased resolutions if used to define the choices available to the parties during a mediation.

Chris Draper

In terms of the Old Riskin Grid, an idealized mediator is typically considered one who operates in the Facilitative-Broad quadrant.<sup>15</sup> These mediators allow the parties to arrive at their own solutions, no matter how unconventional the outcome. In this quadrant, the mediator facilitates an environment that reduces emotion and expands creativity. In these mediations, the mediator can read the room, feel where things are going and enable breakthroughs. Many mediators who succeed in this quadrant are known for their empathy, charisma and likability, which are observable personality traits that empower some to assess mediator quality by “looking at the person’s background, formal mediation training, and biases”.<sup>16</sup> This focus on interpretable human qualities allows many to see mediation as more art than science.

Yet modern research indicates that most effective mediators are successful because of their consistent use of Pull Style communication.<sup>17</sup> Using metrics that define the presence of Pull Style communication, feedback accessible through ODR platforms and NLP AI tools, an automated mediator that conforms to ethical norms is no longer science fiction.

## 7 Automating Mediation

Accepting the latest research on effective negotiation strategy indicates that an unbiased AI mediator that enables the parties to more rapidly arrive at any outcome that is agreeable to all parties maximizes Pull Style communication.<sup>18</sup> Maximizing Pull Style communication can be distilled into the following three major components:

- Building upon areas of agreement,
- Seeking information, and
- Facilitating inclusion.

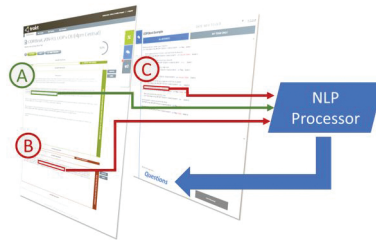
When impasses occur, research indicates that consistently successful mediators routinely employ Pull Style communication strategies of:

- Clarifying aspects of the conflict that could be interpreted in a different manner,
- Asking questions that seek out whether the needs of one party could be built on the needs of another, or
- Engaging each party so they are equitably contributing.

- 15 L. Riskin, ‘Understanding Mediators’ Orientation, Strategies, and Techniques: A Grid for the Perplexed’, *Harvard Negotiation Law Review*, Vol. 1, No. 7, Spring 1996, <https://scholarship.law.ufl.edu/cgi/viewcontent.cgi?article=1684&context=facultypub> (last accessed 7 July 2019).
- 16 C. Currie, ‘Mediating Off the Grid’, *Mediate.Com*, November 2004, [www.mediate.com/articles/currieC4.cfm?nl=64](http://www.mediate.com/articles/currieC4.cfm?nl=64) (last accessed 7 July 2019).
- 17 A. Abramowitz, ‘How Cooperative Negotiators Settle without Upending the Table’, *Design Intelligence*, 20 October 2006, [www.di.net/articles/how\\_cooperative\\_negotiators\\_settle/](http://www.di.net/articles/how_cooperative_negotiators_settle/) (last accessed 7 July 2019).
- 18 A. Abramowitz, *Architects Essentials of Negotiation*, 2nd ed., London, Wiley, 16 March 2009, ISBN-10: 9780470426883.

Each of these actions can be achieved with NLP AI tools. For example, the following displays an example negotiation in Trokt where an NLP processor finds three related items (A, B and C) that conflict both with each other (B conflicts with C) and with an agreed element (B and C conflict with A).

**Figure 5** *Example of Negotiation in Trokt Where an NLP Processor Finds Disagreement*<sup>19</sup>



When the NLP processor identifies these disagreements within related elements, it could be trained to propose discussion questions such as the following:

- [Clarifying] ‘Can you more fully explain what comment C means?’
- [Clarifying] ‘Do you see a relationship between B and C?’
- [Building] ‘Is there a way that B could be incorporated into A?’
- [Building] ‘If C were added into B, could it be incorporated into A?’
- [Engaging] ‘Team 1, could you clarify...’
- [Engaging] ‘Team 2, could you propose...’

Using an NLP processor trained to look for related items that contain enough uncertainty or inconsistencies to indicate conflicting positions would produce an AI mediator that is able to continually question the parties without bias. Unlike systems that may observe A, B or C in the above example and offer alternative suggestions given the context of similar negotiations, focusing on Pull Style questions confined to the facts in the negotiation will avoid the AI mediator from perpetuating the systemic bias inherent in filtering-based tools. Further, while the NLP processor could be built to prioritize questions around related items with the most significant uncertainty or inconsistencies first, it is more likely that the NLP processor would find minor inconsistencies that could be overlooked by a human mediator that is developing patterns of interpretation within the negotiation that are unconsciously biased by past experience.

This design approach will enable true facilitative mediation, which can be augmented by checks to ensure neither party agrees to anything that is counter to law, yet directly avoid the systemic bias inherent in historically calibrated filtering ODR platforms.

<sup>19</sup> C.H. Draper, ‘Trokt AI Design Specification’, Internal Meidh Technologies Document, 16 January 2018.

Chris Draper

Current NLP tools and negotiation communication research indicate that there is a clear, objective path for creating a fully autonomous NLP AI mediator built on Pull Style communication algorithms. However, it is also clear that hesitation within the access to justice and dispute resolution communities will likely require the first AI mediator to gain iterative acceptance. Assuming the need for iterative acceptance, designers may find value mimicking successful autonomous vehicle development pathways when developing a rollout strategy that avoids access to catastrophic failures. Like autonomous vehicles, the role of an AI mediator is to set the boundaries within which users will operate. Like autonomous vehicles, human participants must be effectively engaged to ensure success. And like autonomous vehicles, an unplanned diversion outside of safe operating conditions could result in a catastrophic consequence for the designer or the wider community.

The National Highway Traffic Safety Administration (NHTSA) has identified its roadmap to fully autonomous vehicles based on the Society of Automotive Engineers (SAE) automation levels 0-5.<sup>20</sup> These levels can be translated into ADR practices as follows:

- Zero (0). No automation, which equates to technology-free ADR in the dispute resolution space.
- One (1). Driver assistance, which equates to using AI tools for email or document drafting that suggest more appropriate language.
- Three (3). Conditional automation, which equates to an autonomous NLP AI tool that requires human verification before any output is approved.
- Five (5). Full automation, where an NLP AI will likely employ some form of Strong AI that can operate without oversight.

20 NHTSA, 'Automated Vehicle Safety', [www.nhtsa.gov/technology-innovation/automated-vehicles-safety](http://www.nhtsa.gov/technology-innovation/automated-vehicles-safety) (last accessed 7 July 2019).

**Figure 6** Society of Automotive Engineers (SAE) Automation Levels<sup>21</sup>

0	1	2	3	4	5
<b>No Automation</b>	<b>Driver Assistance</b>	<b>Partial Automation</b>	<b>Conditional Automation</b>	<b>High Automation</b>	<b>Full Automation</b>
Zero autonomy; the driver performs all driving tasks.	Vehicle is controlled by the driver, but some driving assist features may be included in the vehicle design.	Vehicle has combined automated functions, like acceleration and steering, but the driver must remain engaged with the driving task and monitor the environment at all times.	Driver is a necessity, but is not required to monitor the environment. The driver must be ready to take control of the vehicle at all times with notice.	The vehicle is capable of performing all driving functions under certain conditions. The driver may have the option to control the vehicle.	The vehicle is capable of performing all driving functions under all conditions. The driver may have the option to control the vehicle.

Mediators who are managing remote communication by sending documents back and forth via email or collaborating internally on an AI-assisted, cloud-based collaboration platform would therefore be currently operating somewhere between a Level 1 or a Level 2. Assuming all steps are equal, this indicates that the ODR world is likely already 30-50% of the way towards full automation. Yet with all technological innovations, it is more likely that the final three steps will come upon us like a dam breaking, meaning fully autonomous NLP AI tools will be ready to be operated sooner than many in the ADR, ODR and access to justice industries may be ready to expect. To ensure facilitative ODR does more than accelerate the adoption of inequitable precedents; now is the time to implement a set of mediation standards that reach past the current filtering problems that typically define court-based ODR.

## 8 Conclusions

The courtroom automation efforts that are implementing ODR tools that depend on efficient filtering have woken up the access to justice and dispute resolution communities to the opportunities and dangers of AI. Systems that use past out-

<sup>21</sup> *Ibid.*

Chris Draper

comes to calibrate an AI system's future decision processes will accelerate the adoption of the systemic inequities that are rife within the current judicial system. Yet a design focus on Pull Style communication strategies and techniques offers a pathway for developing efficient, effective and unbiased AI mediator tools that can be used to assist, share or eliminate human mediator workload. Given how quickly the industry is expected to move from its current level of automation to full mediator automation, now is the time for transparent discussions on how we expect these tools to behave. If our communities are committed to understanding the modern ghost in the machine, we can remake tools that are currently on a path towards accelerating inequity so they can help rewrite the old concepts of law and order into solutions that meet the needs of our communities.