Artificial Intelligence and the Future of Online Dispute Resolution

Dave Orr and Colin Rule

One concept that has seized the popular imagination is the idea of the digital judge. There is something intuitively appealing about the concept that one day our unruly, chaotic human disputes will be resolved by the cool, all-knowing rationality of a fair and impartial electronic decision maker. While the concept may be enticing, this leap from human-powered justice to electronic justice is a pretty big one. Much like the concept of self-driving cars or watches we can talk into, many people seem to have concluded that this future is inevitable, even when we don't yet have the technology that could make it come to pass. Right now we're just biding time, waiting for the future to arrive.

There are several reasons why we feel the arrival of the digital judge is inevitable. First, we humans generate billions of disputes each year, soon to be tens of billions. This growth shows no signs of stopping. We cannot help ourselves; we love to fight with each other. Despite this love of fighting, the idea that current, inefficient, human-based resolution processes could resolve all these disputes strains credulity. Faith in our very ability to be fair and impartial arbiters weakens under this strain, and it is undermined even further by what we continue to learn about how our brains work. Alongside these developments, computers continue to become more powerful and more deeply integrated into our everyday lives. It stands to reason, then, that if current trends continue, computers will one day be better at fairly resolving our disputes than we are.

Considering this, one thing becomes clear: if computers are going to resolve our disputes, they are going to go about it in a very different way than we have up until now.

I. Technology, Dispute Resolution, and the Fourth Party

Online dispute resolution (ODR) is the use of information and communication technology to help people prevent and resolve disputes. ODR, like its offline sibling *alternative dispute resolution* (ADR), is characterized by its extrajudicial nature. In a sense, dispute resolution is defined by what it is not: it is not a legal process. Any resolution outside of the courts is dispute resolution. If you and your counterparty decide to resolve your dispute by consulting tarot cards, that is alternative dispute resolution. If you decide to resolve your dispute with a game of checkers, that is also alternative dispute resolution. However, if you decide to resolve your dispute with a game of *online* checkers, that is online dispute resolution. Either way, in the dispute resolution world, we paint with a pretty big palette.

As ODR has developed over the past 20 years, a few core concepts have emerged. One of the most foundational concepts is that of the "fourth party". Originally introduced by Ethan Katsh and Janet Rifkin in their book *Online Dispute Resolution*,[1] the fourth party describes technology as another party sitting at the table, alongside party one and party two (the disputants) and the third party (the neutral human, such as a mediator or arbitrator). You may be forgiven for picturing the fourth party as a friendly robot sitting next to you at the negotiating table and smiling patiently. Bear in mind, though, that this fourth party could just as easily be a black

[1] Ethan Katsh & Janet Rifkin, *Online Dispute Resolution* (Jossey-Bass, 2001).

cylinder sitting on the table - *a la* Amazon Echo - or just software floating somewhere in the cloud. The form of the fourth party is irrelevant to the function the fourth party provides.

The fourth party can play many different roles in a dispute. In most current ODR processes, the fourth party is largely administrative, handling tasks like case filing, reporting on statistics, sharing data, and facilitating communications. We ask our friendly fourth party robot to take notes, or to dial in someone who could not join us at the table in person. But it is obvious to those of us in the ODR field that the fourth party is capable of much more. While we humans pretty much work the way we always have, with our cognitive biases and attribution errors, computers are getting more powerful all the time. It is inevitable that at some point we will ask our fourth party robot to help us resolve our issues, or maybe even to just handle it for us outright. The fourth party is just getting started.

II. Getting Used to the Machines

There was a time when technology was perceived as very dehumanizing. Dispute resolvers in particular resisted the idea that algorithms had any useful role to play in helping disputants find solutions to their disagreements. But technology has become much more accessible and integrated into our lives, and we now use technology in ways we never would have considered ten years ago. People take to the internet to find their spouses, to find information on where to go to church, to choose the best school to send their kids, and even to seek out a cardiac surgeon. The younger generation is even more comfortable: they ask each other to prom, break up over Twitter, and Snapchat their friends embarrassing pictures from last night's party.

Individuals have come to trust information presented to them by an algorithm more than they trust information presented by a human. While this might seem initially jarring, upon reflection, it makes some sense. If you are thinking about getting a divorce, you may want to consult a lawyer to learn about your rights and the required steps. Perhaps in the consultation with that lawyer, you feel they are judging you in some way – maybe for your age, or for your ethnicity, or even for your perceived ability to pay. Maybe you suspect that the lawyer is wondering whether the divorce is your fault, or is tailoring the information he or she presents to you in order for you to pick a resolution process that the lawyer feels is more appropriate in your particular situation. That feeling can be very uncomfortable.

Now think about an algorithmic consultation. You go to Google and type "divorce". The search returns hundreds of millions of results, and you scan through the first twenty-five to see if any appear to be on target. You select, somewhat at random, a guide published by a legal service bureau a few counties away from you. This online guide was clearly not created specifically for you – it was put online several years ago, long before you ever thought you would need to consult it. None of your personally identifiable information is required to navigate the guide. You can answer high level questions about your situation (*e.g.*, do you have kids, are you both employed) without providing your name. After six or seven minutes of navigation and simple questions, the guide shares its conclusions about the likely steps that would be involved in your divorce. If the result seems questionable, you can merely reload the homepage and start again, perhaps providing different answers to see how your changes alter the final results. In any event, this algorithmic process does not judge you on the basis of your race, sex, age, income, or other characteristics – largely because it knows nothing other than what you tell it. It stands to reason that people might be more comfortable using assessment tools such as this one when they are trying to get their questions answered. In addition, that algorithm is probably free, while a lawyer will probably charge an hourly rate for the same service.

III. The Rise of Artificial Intelligence

Technology is likely to alter many areas of professional services, from financial planning to medical care. But in the justice sector, this development may prove particularly significant. Government has an interest in the consistent resolution of disputes, and to that end, government funds the courts. But it is unlikely that the government will be the sole provider of algorithms used in these ways. Just as the internet has weakened the role of the public sector in many areas of the economy (*e.g.*, bitcoin has made financial transactions stateless and invisible to regulators), it may also weaken the role of the public sector in providing justice.

A shorthand for the expansion of technology into these realms formerly dominated by humans is the term *artificial intelligence*, or AI. People often envision AI working the way humans work, perhaps taking the form of a humanoid robot in the front of a courtroom, wearing a powdered wig on his metal head and wielding a gavel in his little robot hand. That image may be drawn more from old episodes of *The Jetsons* than from technological necessity, but sometimes there is value in matching people's expectations. If that form is more satisfying to people, it certainly is doable. In reality, though, the action in AI takes place in software, no powdered wigs necessary.

AI uses software algorithms to tackle complex tasks that have been traditionally handled by non-artificial intelligences (*i.e.*, us, the humans). Humans have their own ways of understanding problems and devising solutions. AI also has to understand problems and devise solutions, so that it can deliver outcomes equivalent to, or better than, human devised outcomes. But algorithmic intelligence doesn't go about devising those outcomes in the same way as human intelligence would.

We've all heard about IBM's Watson winning *Jeopardy!* over the top human players in the world. Many of us might presume that Watson works like an electronic human brain, mimicking the same types of connections that happen in the human players' brains during the game. But that isn't the way Watson is programmed to operate. As Alex Trebek is reading off each word of the question, Watson is guessing what the question is getting at, and instantly generating thousands of possible responses to the possible question. Watson is scoring all of those possible responses in real time, estimating the likelihood that each one is the right answer. As soon as Watson finds an answer with the highest likelihood of both 1) the question being the right question, and 2) the answer being the correct answer to that question, Watson buzzes in. The other human players are trying to make connections in their brain that generate the one best answer, but Watson is generating thousands upon thousands of answers and scoring them all to see which one is best. This is similar to the way computers win chess matches: they evaluate all possible moves one move out, two moves out, and three moves out; score them all; and then decide which move is best in each situation. This is fundamentally different from the way a human plays chess or plays *Jeopardy!*, but the result is equivalent to or even better than a human's performance.

When an AI is first created, it is a blank slate hungry to learn. But as we can see from the above examples, an AI learns in a very particular way. It learns by looking at data, and this data must be structured into a format that the AI can make sense of. The AI can then look at this data in order to formulate some observations, but to train an AI to make these observations, you must always first provide the AI a corpus of data.

For example, imagine an AI is asked to decide the appropriate penalty payment owed by a business for inappropriately sharing a consumer's private information. Maybe there is a large database of prior cases that contains more than ten thousand decisions made by customer service representatives about penalty payments. The details of each of these violations (such as severity, scope, and type of information shared) are stored in the database. The algorithm then crawls through all of the cases and creates a set of rules that correlate the decision rendered in each case to the details of each case. With this setup, when a new case is presented to the AI, it will consult the rules it already created when it learned from the corpus, and it will then make a determination as to the appropriate payment amount.

This algorithm is built from determinations originally made by the customer service representatives. Let's say the reps were very skilled at making their determinations, but were still wrong about 10% of the time. Because the algorithm trains itself based on these decisions, the AI cannot make the correct decision more than 90% of the time. The algorithm cannot use the data in the corpus to train itself to be *better* than the data set it was presented. But the bigger the corpus, the more specific the AI will be in crafting rules, and that will enable the AI to get ever closer to that 90% accuracy level.

Sometimes a corpus of data might not exist around a particular decision type. For example, imagine there is a need to decide if a certain online review is specific enough for inclusion on a hotel rating website. No database exists that contains prior evaluations of reviews to determine if they meet the standards in question. But perhaps the hotel rating website starts a crowdsourced process to evaluate reviews. Members of the website are repeatedly asked if a particular review is specific enough for inclusion. Every time they log in they get another review to evaluate. Maybe customer service reps also decide some cases as well, in addition to the users. Slowly but surely, website members and customer service reps would generate a corpus of data. As each decision is rendered, the AI could be watching and learning from each new case. Again, maybe the users only get it right about 90% of the time, but by observing enough of these evaluations, and by capturing all of the outcomes from the crowdsourced process in a structured database, the AI algorithm could train itself what to look for, and eventually be able to make decisions about future online reviews at a similar level of accuracy. At this point, the human-powered crowdsourced decisions could taper off, and the AI algorithm could increasingly take over.

When AIs come up with rules, it may seem like magic. You might even want to open the hood and see just what these miraculous rules are, so that you can leverage them in your own decision-making. Don’t bother. Most humans cannot make heads or tails of the rules AIs glean from a large corpus of data. For example, an AI may decide that a review that has the word “actually” within eight words of the word “budget” is likely to be a trustworthy review. Now why is that? Our simple little human brains might not be able to come up with a good explanation as to why that may be true. But the AI has found a pattern, and that pattern may have truth undergirding it that a human is not able to comprehend. In fact, if you look at most rules generated by AIs, they appear to us humans as gobbledygook. But that is only because humans think like humans, and AIs think like computers. There may be insight in those rules that we are simply unable to understand. As they say, the proof is in the pudding, and if the output is high quality, then the logic generated by AIs is quality, even if to humans it doesn’t seem all that logical.

IV. Building the Corpus

The challenge is not necessarily to think about how to train an AI to decide a dispute. As we’ve already described, we know how to program an AI so that it can take on that task. The real challenge is, how do you categorize the world’s resolution information into a format an AI can make sense of, and not only make sense of, but learn from?

There is no shortage of raw data in the world. There are lots of court decisions that we could give an AI to read, for example. There are also many companies out there trying to make sense of court cases via AI. The problem lies in finding ways for AI to process this data. Currently it is very difficult to do. There is a lot of structure to the law, but it is not the kind of structure that can easily help an algorithm learn and identify patterns and rules. We are still a long way away from giving an AI Lexis-Nexis access and then asking it to serve on the Supreme Court.

So what do we do? If we want to train AIs to be better decision makers, we need to build data sets. Since so many cases are now being decided on ODR platforms, one task AIs could take on in the near term would be to help build these data sets through case classification. Humans would negotiate, mediate, and arbitrate new cases, and AIs would review the outcomes and structure the data they generate in real time. This would give us a good head start on building a large corpus we could use to train future AI algorithms. AIs are very good at labeling data and storing it in a structured way that will make sense to future algorithmic analysis. If an AI labels and classifies millions of traffic court decisions in real time, for example, then we can open that database to other algorithms that could then use that data to educate themselves about traffic cases. This could potentially teach all those algorithms how to accurately decide traffic court cases moving forward. It's a long way from the Supreme Court, but it's a start.

This is an important point, and an important limitation to consider. An AI must focus on similar baskets of cases. It is very difficult for an algorithm to get a database of many different kinds of cases (*e.g.*, workplace, traffic, divorce) and then somehow glean rules that could make sense out of any possible new case. Specialization into specific case types (*e.g.*, traffic) is very important for accuracy in rules. General decision-making systems (humans) still need to be able to determine the classification of each new case, and then apply the rules relevant to that specific case type. AI is not there yet, but perhaps one day a team of AIs will work to resolve cases, with the first AI routing each incoming case to the appropriate queue, and a second AI determining the appropriate outcome for cases of a particular type.

V. Changing the Way We Think About Justice

The techniques we are describing are feasible today. But if that's the case, where are all of the algorithmic judges? The truth is that they are out there, silently churning away, but currently they are primarily focused on answering relatively simple data-based questions.

The reason for this is that AI algorithms are still not very good at making sense of unstructured data. For example, if we were to show the transcript of a negotiation session to an AI and ask that AI to suggest a fair resolution, that would require some pretty advanced capacity on the part of the AI. In the near term, the speech transcription of the session is being solved, so the AI can probably learn the words said in the session. But words are only part of what is communicated in a negotiation. Identifying the truly important points of disagreement in a dispute, and comprehending the subtexts and assumptions behind each of those points, is much harder. Teaching an AI to contextualize unstructured communication may be possible in 10-20 years, but at the current moment AI may get just as confused by legalese as a layperson.

What breakthroughs are required to help AIs get over that hump? How could an AI gather more understanding to fill in the blanks in a negotiation? Maybe AIs can be taught to ask the disputants questions, the same way a judge or a mediator might, in order to get at more subtle points of meaning. Perhaps the AI could educate itself by reading the internet, or looking through case databases to try to learn from similar matters. The AI could then bring conclusions drawn from other cases into each new conversation, which could help it parse points of confusion without constantly asking the parties to explain what they mean by each comment they contribute.

One way we could make it easier for algorithms to resolve our disputes is to structure our negotiations into questions that are more easily answerable by computers. For example, instead of asking an algorithm to simply issue a decision from scratch in a disagreement, perhaps the two parties in a disagreement could be asked to put forward their last, best offer, and the algorithm would be asked which of the final offers is more appropriate. In this design, the algorithm would conduct research in databases around the world, return a result, and then see which of the proposals is the closest to its template resolution. The parties would also have an incentive to be as reasonable as possible in putting their offers on the table, because they would want the AI to pick their suggested resolution over the other party's proposed solution. This kind of technology-assisted final offer arbitration could be a shortcut to AI-powered resolutions, because this design plays to algorithmic strengths and avoids difficult, more nuanced questions that might trip it up. It also avoids the possibility that the AI really gets it wrong and delivers a resolution that is wholly unjustified, frustrating both of the parties.

There are intermediate steps on the road to the digital judge. AIs do not have to serve as the final decision maker right out of the box. AIs could start out by evaluating cases and coming up with suggested resolutions that human decision makers might consult on an advisory basis. Parties could also run their cases by an algorithm in advance of a human-powered arbitration to see what resolution the algorithm might consider fair. Even the best arbitrator can only keep a couple hundred case outcomes in their mind, but an algorithm can consult millions or tens of millions of cases and factor all of that information into its suggested resolution. Consulting AIs in this way could not only help to improve the quality of AIs, but also increase confidence in the ability of AIs to render trustworthy decisions. Once the AI has proven itself effective – perhaps after consulting on millions of cases – then it could be put into the final decision-making role.

VI. Deciding What AIs Can Consider

AIs act very differently from people, but these differences may actually be beneficial. AIs can be programmed in a way that makes them more "fair", by ignoring information that system designers and programmers deem to be outside the scope of the question at hand. For example, you can never be sure whether your jury was swayed by some unforeseen factor, like your accent or your hemline. The jury may not be sure themselves as to why they feel compelled to decide your case one way or the other, but a computer algorithm can not only be explicitly instructed to ignore certain factors (*e.g.*, accent, hemline), but it can also be prevented from even knowing those bits of data in the first place. There is no way for a jury to ignore such factors, not even after explicit instructions from the judge not to pay attention. There is a surefire way, though, to prevent the AI from knowing them.

This leads to some interesting design choices – and complex ethical and moral ramifications – for building dispute resolution AI. For example, computers have gotten very good at reading human facial expressions. Is it reasonable for a computer to closely watch a disputant explain their actions, and then to determine based on the observed facial expressions whether the explanation is a lie? What if the computer could conduct an MRI on the disputant as they offer their explanation, and from that MRI provide certifiable evidence that the statement is a lie? Should that information be factored into the computer's decision-making process, or should the AI be forbidden from considering it? It is up to the AI's programmer to determine if that information is relevant, as well as whether the algorithm will even be capable of gathering this kind of data during the dispute. There may be a certain ick factor in giving computers so much visibility into things that we as humans cannot perceive ourselves. But we may conclude that the accuracy and accountability that comes from these new capabilities may outweigh the ick factor, and our instinctual resistance may ease over time.

On the other side of the coin, AI systems might make egregious mistakes that humans would never make. This may, however, be due to their systems designers failing to integrate all of the information required to avoid such mistakes. For example, Google's self-driving car follows the explicit laws on the books that regulate driving, but it does not follow the *implicit* rules that so often conflict with the laws on the books.[2] A human understands both these sets of rules, and appropriately contextualizes them in real time. A machine might not know both sets of rules unless there is some way to integrate them into the algorithm. To picture the problems of this lack of context, imagine a human driver seething behind a row of Google cars all driving the

---

[2] Matt Richtel and Conor Dougherty, "Google's Driverless Cars Run Into Problem: Cars With Drivers," *New York Times*, September 1st, 2015, last accessed November 28, 2017, https://www.nytimes.com/2015/09/02/technology/personaltech/google-says-its-not-the-driverless-cars-fault-its-other-drivers.html?_r=1.

exact speed limit.Sometimes AIs may make decisions that seem odd or ill-advised to a human observer, and it can be very hard to understand the reasoning behind an AI's decisions. By carefully deciding the information AIs are given, and by working out the kinds of decisions AIs are allowed to make with that information, all of these kinks can eventually be worked out, and AIs can gradually become more integrated into the decision-making process.