Artificial Negotiators

Introduction

The idea of robots displacing human labor has captivated the public imagination for decades.[1] Robots and other computerized systems have in fact displaced many human workers and revolutionized many industries.[2] Recently, the rapid development of autonomous vehicles has brought back to the fore the public anxiety over artificial employees.[3] The expectation of displacement usually focuses on jobs characterized as routine or manual, or both.[4] Those in non-routine, cognitive occupations like attorneys have typically been viewed as least susceptible to replacement by computerized systems,[5] but this attitude is changing even amongst lawyers themselves.[6] Some major firms already have programs that are exploring the use of systems based on IBM Watson to replace certain associate functions.[7]

---

[1] Consider the proliferation of famous artificial beings in popular culture that were designed to fill occupational roles: "Robby the Robot," FORBIDDEN PLANET (1956); "Rosie," *The Jetsons* (1962); "Robot B-9," *Lost in Space* (1965); "HAL," 2001: A SPACE ODYSSEY (1968); "R2-D2" & "C-3PO," STAR WARS (1977); "Marvin the Paranoid Android," DOUGLAS ADAMS, THE HITCHHIKER'S GUIDE TO THE GALAXY (1979), "Bishop," ALIENS (1986), "Data," *Star Trek: The Next Generation* (1987); "Bender," *Futurama* (1999); "WALL-E" & "EVE," WALL-E (2008).

[2] *See, e.g.*, Jane Wakefield, *Foxconn Replaces '60,000 Factory Workers with Robots'*, BBC NEWS (May 25, 2016), http://www.bbc.com/news/technology-36376966/; Rebecca L. Weber, *The Travel Agent is Dying, But It's Not Yet Dead*, CNN, (Oct. 10, 2013, 6:20 AM), http://www.cnn.com/2013/10/03/travel/travel-agent-survival/; Andrew McAfee, *Manufacturing Jobs and the Rise of the Machines*, HARV. BUS. REV. (Jan. 29, 2013), https://hbr.org/2013/01/manufacturing-jobs-and-the-ris/.

[3] *See, e.g.*, Ryan Petersen, *The Driverless Truck is Coming, and It's Going to Automate Millions of Jobs*, TECHCRUNCH (Apr. 25, 2016), https://techcrunch.com/2016/04/25/the-driverless-truck-is-coming-and-its-going-to-automate-millions-of-jobs/; Jason Del Rey, *Uber CEO: Self-Driving Cars Are the Future, Drivers Are Not*, RECODE (May 28, 2014, 2:14 PM), http://www.recode.net/2014/5/28/11627344/uber-ceo-self-driving-cars-are-the-future-drivers-are-not-2/.

[4] *See* David H. Autor, Frank Levy & Richard J. Murnane, *The Skill Content of Recent Technological Change: An Empirical Evaluation*, 118 Q.J. ECON. 1279, 1282–1286 (2003); Andrew Flowers, *The Shift from Low-Wage Worker to Robot Worker*, FIVETHIRTYEIGHT (Mar. 25, 2014, 7:19 AM), https://fivethirtyeight.com/features/your-new-fast-food-worker-a-robot/.

[5] *See* Flowers, *supra* note 4.

[6] *See* David Kravets, *Law Firm Bosses Envision Watson-type Computers Replacing Young Lawyers*, ARSTECHNICA (Oct. 26, 2015, 1:06 PM), https://arstechnica.com/tech-policy/2015/10/law-firm-bosses-envision-watson-type-computers-replacing-young-lawyers/; Julie Triedman, AmLaw Daily, *Computer Vs. Lawyer? Many Firm Leaders Expect Computers to Win*, THE AMERICAN LAWYER (Oct 24, 2015), http://www.americanlawyer.com/home/id=1202740662236?slreturn=20170201191139/.

[7] For example, Dentons and Latham & Watkins are using ROSS, a legal research tool based developed in a collaboration between IBM and the University of Toronto. *See, e.g.,* Triedman, *supra* note 6; Jeff Gray, *U of T*

Artificial Negotiators

All lawyers, and other workers in primarily non-routine, cognitive occupations, still perform numerous routine tasks many of which have already been partially shifted to technology. But can an artificial intelligence system completely replace a lawyer in the complex, dynamic, and multi-faceted process of a negotiation? This essay will consider how an algorithmic system may in fact be able to do just that, and how such an artificial negotiator functionally compares with a human lawyer. Part I will define the sophistication of the artificial negotiator that will be considered. Part II will describe the approach an artificial negotiator would take compared with an experienced human attorney. This part will also explore which aspects of negotiation the system could excel at, as well as those areas where it may struggle. The discussion will be organized based on the seven elements of negotiation as taught by the Harvard Law School Project on Negotiation: interests, options, legitimacy, alternatives, commitments, communication, and relationship.[8]

I. Defining the Artificial Negotiator

It is easy to dismiss the idea of computers performing complex processes like negotiation for a couple reasons. First, a task like a negotiation is so unpredictable that it seems antithetical that any computer programmer could code a system to respond to the unlimited possibilities. Second, the skills that lawyers and other negotiators are commonly thought to employ – things like creativity and judgment – are not skills that any known computer is thought to possess. Richard and Daniel Susskind have characterized this kind of thinking as the "AI Fallacy" – the notion that computers need to replicate human thought patterns in order to equal or surpass human abilities.[9]

Students' Artificially Intelligent Robot Signs with Dentons Law Firm, THE GLOBE AND MAIL (Aug. 10, 2015, 9:09 AM), http://www.theglobeandmail.com/report-on-business/industry-news/the-law-page/u-of-t-students-artificially-intelligent-robot-signs-with-dentons-law-firm/article25898779/.

[8] For a description of these elements, see, e.g., ROBERT H. MNOOKIN, BEYOND WINNING: NEGOTIATING TO CREATE VALUES IN DEALS IN DISPUTES (2004).

[9] RICHARD SUSSKIND & DANIEL SUSSKIND, THE FUTURE OF THE PROFESSIONS 45 (2015) ("[S]ystems of today are increasingly out-performing human experts, not by copying high-performing people but by exploiting the distinctive capabilities of new technologies, such as massive data-storage capacity and brute-force processing.").

The android-style artificial intelligence of science fiction may one day exist, but before that, other highly effective forms of computer intelligence will emerge that employ different analytical approaches.[10]

Individual computing systems with the processing power equivalent to an adult human brain (approximately 10^{16} computations per second) are predicted to be widely available by 2020.[11] At similar advancement rates, by 2050 that processing power will eclipse the collective computational capability of all humans on Earth – in *each* machine.[12] That raw analytical ability may not be able to think creatively, but it can successfully attack problems in ways that humans cannot. One way is via complex analysis of very large data sets – so-called "Big Data."[13] It was recently estimated that the world is producing as much data every two days as has been produced in most of human history.[14] The supercomputers of the future will be able to access and process some or all of this extensive data reserve in ways that cannot be fully anticipated to discover correlations and abstractions of incredible complexity.[15] Those discovered patterns can in turn be used to probabilistically solve problems without conscious comprehension of the problems by the analyzing machine.[16] This is similar to how IBM Watson beat record-setting champions and very smart humans Ken Jennings and Brad Rutter on *Jeopardy!*.[17] Watson employs natural language processing to parse textual inputs,[18] but it doesn't consciously comprehend the question that are

[10] *See id.* at 45, 278 ([W]hether machines will replace human professionals is not about the capacity of systems to perform tasks as people do. It is about whether systems can outperform human beings—full stop.").

[11] *Id.* at 157 (citing RAY KURZWEIL, THE SINGULARITY IS NEAR 7-8 (2005)).

[12] *See id.*

[13] The term Big Data in this context generally refers to data analytics on large data sets, not the large data sets themselves; it sometimes also refers to the data analytics industry. *See id.* at 161.

[14] *See id*. at 161, 161 n.38 (based on observations in 2010 by Google CEO Eric Schmidt that 3 exabytes of data are produced each day, while only about 5 exabytes had been generated before 2003).

[15] *See id.* at 163–64.

[16] *See id.* at 185–87.

[17] *See* Jo Best, *IBM Watson: The Inside Story of How the Jeopardy-Winning Supercomputer Was Born, and What It Wants to Do Next*, TECHREPUBLIC (September 9, 2013, 8:45 AM), http://www.techrepublic.com/meet-the-team/uk/jo-best/?q=&o=2&t=&m=10&topic=&d=5&a=.

[18] *See id.*

asked of it. The version that appeared on *Jeopardy!* performed searches of encyclopedic knowledge based on the words and structure of the clues, and then built a probabilistic model of potential answers.[19] Watson didn't mindfully answer questions, so much as it made a statistical determination of what answer the question was most likely designed to target.[20]

An artificial negotiator would approach negotiation with a similar method of processing a large number of inputs and using probabilistic decision trees to determine its responses or decide on the best action. In this essay, I am assuming a system that has the processing power of a small team of humans. The system has natural language processing abilities that surpass IBM Watson, and can correctly parse even the most complex spoken and written sentences. It can read a contract and convert all the contained provisions into code. It can evaluate how useful a provision is for a specific client goal. It can read a legal opinion and follow the argument. It can prognosticate and even write a legal opinion based on hypothetical facts; but it won't be eloquent or use creative examples. It can draft a contract, and can do so in multiple ways to achieve a desired outcome. This should not all be confused for creativity or imagination, however; the negotiator can't write poetry or tell an original joke.

The negotiator also has access to any information that any person or entity could access for free or with a reasonable fee or subscription, including: government records, news sources, data from financial markets, social media, digitized libraries, all public legal documents, published scientific and social science literature, and any document available on the internet. Many negotiated agreements and the details surrounding their genesis and aftereffects may not be publically accessible. But the algorithm would have access to those that were public, for example, merger agreements of public companies or employment agreements for certain union and

[19] *See id.*
[20] *See id.*

government workers.[21] It would be able to build from the raw and scattered documents, a cohesive database of negotiations to use as a comparison tool, a source for options, and data for predictions.

Lastly, for the purposes of this essay I am assuming that the artificial negotiator is an agent representing a client, not negotiating on its own behalf. The envisioned counterparty is also represented by an agent, be it human or artificial. The artificial negotiators, as I am considering them, do not have their own interests and they themselves do not experience principal-agent tension.

II. How Well Can an Artificial Negotiator Perform?

1. Interests

Interests are challenging to put in to words, even for the principals in a negotiation and even when they are humans. Lawyers and other agents have an added hurdle of cognizing the principals' interests secondhand. Artificial negotiators would also have to perform this task, but the more immediate consideration is how a computer can codify a client's interests and function with a goal of meeting those interests.

Interests that are essentially monetary are easy for a computer to understand and an algorithm can readily understand that more money for a seller and less for a buyer are generally desirable. Even when there are different categories of monetary exchanges, exchanges with varied timing, or exchanges with varying conditions and contingencies, an algorithm would be able to calculate present values and further discount or augment as necessary to account for probabilities

[21] Depending on what entity creates or operates an artificial negotiator it may also have access to private documents. For example, if the negotiator is associated with a data platform like Google, it may be able to process all users' Gmails to analyze more informal negotiations. This ability may be restricted by privacy law and the terms of service for users providing the data to such platforms.

of various triggering events. For such numerical matters, an algorithm would function similarly to an accountant or financial analyst with a sophisticated Excel model.

But there are other interests that are either too attenuated from finances to be meaningfully quantified or are genuinely unquantifiable. These include reputational concerns, relationship building, flexibility, concerns over specific invaluable or sentimental goods, and many more. In the broadest sense, some of these interests may be simply happiness or satisfaction on the part of the principal. Even an artificial negotiator with perfect natural language comprehension would not be able to "understand" these interests in a human sense. For example, a computer may be able to parse a job applicant's reputational interest expressed as, "I don't want to push for so much salary that my future co-workers become resentful." But the computer itself has never experienced resentment.

So how can it attempt to negotiate in a way that is cautious about this risk? One possible way is to turn to Big Data. The artificial negotiator could turn to millions of social media interactions to find interpersonal exchanges where people expressed the language of resentment. Perhaps the exponentially expanding stores of data globally will provide a catalog from which the artificial negotiator can mine more precisely for transactions that led to resentment and compare them to cases where the aftermath was more harmonious. From this analysis, the artificial negotiator can derive an *interest function* for how aggressive salary negotiation relates to risk of resentment. Maybe the algorithm concludes that asking 15% over market is seen as self-advocacy, but pushing for 20% begins to be viewed as selfish.

An artificial negotiator could thus build many *interest functions* that begin to relate specific positions and possible negotiation provisions with its principal's interests. The system also needs a way to make trade-offs between interests: a way to prioritize. As with interests, it is not straightforward for a principal to directly list their interests in order. Such a ranking may be too simplistic to capture such issues as the interests being interdependent or the priorities shifting over a spectrum of negotiated outcomes. Furthermore, a principal may not always be fully sure of what is most important or how to communicate complex relationships between interests. But as a starting place the algorithm can use stated rankings and levels of importance to construct a *priority map*.

A lawyer's best tool to develop a greater understanding of interests and the relationships between them is to ask questions to gather more information. This is likely the same for an artificial negotiator. Based on stated interests and priorities the system can begin to build an *agreement model*. It can base this model on pre-coded *interest functions*, both monetary and non-monetary, and also newly constructed *interest functions* for unfamiliar interests by using data-driven learning as described for the reputational example above. But this initial model is just a guess, the algorithm can refine the model by asking the principal questions: Would you take $100,000, with no options? Would you prefer $90,000, with options valued at $5,000 to $20,000? Would you trade $5,000 in salary for a fifth week of vacation? The market rate is around $80,000, would you accept that, or is it important to you to get to six-figures? The algorithm can then refine the *interest functions* and *priority map* to improve the *agreement model* until the system is quite adept at guessing the scope of negotiated agreements that a principal will and will not reject.

One vulnerability of the artificial negotiator, though this applies to human agents too, is if the principal is uncooperative or not forthcoming. The *agreement model* can only be as accurate as the principal is honest and as precise as the principal is willing to help refine it. Another weakness of the algorithm is dealing with irrational interests and priorities. People, even those acting on behalf of corporations, do not always choose to maximize economic benefit or personal satisfication, even when such outcomes are identifiable. Irrationality, and other "human" behaviors like inconsistency, uncertainty, and confusion may manifest in an unstable and ambiguous *agreement model*. But the model doesn't have to be perfect, it can continue to be refined throughout the negotiation and has interplay with the other elements as I will discuss below.

2. Options

The artificial negotiator's handling of interests that I have described is really just a mechanical analog of how a human agent already tries to explore and balance their principal's needs. Options are the first place that a computer's capabilities truly begin to exceed those of its human counterparts. Although we don't typically think of computers as creative, they have enormous raw "brainstorming" ability. Deep Blue beat Garry Kasparov not by making "creative" chess moves, but by conceiving and evaluating millions of times more moves than Kasparov could ponder.[22] So too, an artificial negotiator can evaluate millions of combinations of assets, contract terms, conditions, contingencies, timing elements, and other instruments or promises that the parties may exchange. The algorithm could use data from any public or otherwise accessible past transactions to build a library of numerous specific examples of assets, terms, etc., from which to deliberately or randomly generate options.

The difficulty for any negotiator comes in how to compare the value of various options or sets of options. The algorithm I have described can compare options using the *agreement model* as test of how each option may satisfy their client's interests and also as a proxy for how likely their client is to accept or veto a certain option. Though I have been focusing so far only on one side of the table, the algorithm could also build a model for the counterparty's interests and priorities. This *counterparty model* could then be used to evaluate the likelihood that a given option

[22] *See* IBM, *Kasparov Vs Deep Blue: A Contrast in Styles – Dissimilar Processes Yield Similar Conclusions*, IBM (1997), https://www.research.ibm.com/deepblue/meet/html/d.2.shtml (comparing Deep Blue's ability to process 200 million chess positions per second to Kasparov's three per second).

satisfies the other side's interests and whether the opposing principal is likely to agree to a given proposal.

Beyond merely testing for the chance of agreement a human negotiator must rely on expertise and past experience to ascertain the various risks of options. An algorithm can similarly rely on expertise and past experience by analyzing the universe of historical data. For example, an algorithm programmed with the basics of contract law can assign a risk value to a particular provision based on certain estimates, such as the likelihood of breach and the chance of obtaining a favorable settlement or judgment. The algorithm can consult any available data on the outcome of historical contracts and contract litigation to empirically derive these probabilities. Through such calculations, the algorithm could rank options with various provisions even without "understanding," in the sense of a conscious human, the implications of those agreements. If the approach is purely data-driven, limited to identifying correlations and estimating probabilities, the algorithm could compare options without even *trying* to cognize what the contract language means. Depending on the details of the pending negotiation, such data analyses could be narrowed to find the preferred terms based on specifics of the parties; for example, refining by the industries, sizes, and locations of two companies in the context of a merger. Human negotiators may come up with similar terms, just because they are consciously following industry practice.

3. Legitimacy

The use of objective criteria is another instance where an artificial negotiator could outpace a real person. Again, this is a result of data access and processing capability. A computerized system should be able to quickly assemble a thorough database of prior comparable transactions and sort through it to eliminate irrelevant benchmarks. It can also index relevant regulations, statutes, and law that may limit or guide negotiated terms. Take for example a job offer negotiation:

the algorithm could find thousands of similar employment contracts and categorize them based on the numerous attributes of individual offerees. Then in constructing an offer for a particular employee it could provide, along with the offer itself, a thorough report of the ranges and averages for salaries, bonuses, options, and other terms based on that employees' traits (e.g., position, experience, education, cost of living).

This basic approach to justifying job offers is already practiced by human human resource departments, but the artificial negotiator would be drawing from a much broader data source. The algorithm could also provide more detail and transparency in reporting to the other side how a particular piece of objective criteria was derived from the data. This also relates to another aspect of legitimacy, which is how legitimacy influences the strategy of making offers and counter-offers. An algorithm could be tuned, based on the indicated aggressiveness of a client, to use more heavily supported or more thinly constructed justifications for a particular anchoring offer.

One thing an algorithm may find challenging is how to accommodate objective criteria from the opponent that conflict with the algorithms own internal calculation of a reasonable position. This may be most interesting in the case where two algorithmic negotiators are battling each other in a war of legitimacy, with essentially the same access to data and the same analytical power. The two algorithms would be adept at realizing how they each have selectively used the available raw data. It may also mean that legitimacy in such negotiations becomes a less emphasized element given the more equal information access, though it will likely still matter behind-the-table in convincing principals than an offer is reasonable.

4. Alternatives

A best alternative to a negotiated agreement (BATNA)[23] or any other alternative is just a non-negotiated option, so the advantages of algorithms to evaluate more options than a human would apply to alternatives as well. An artificial negotiator may be able to identify many more alternatives and more complex alternatives than a human. The client's BATNA would thus be optimized greatly versus the handful of walk-away plans that a human negotiator might seriously evaluate. The artificial negotiator can construct myriad combinations of actions and transactions that might make up a given alternative, using Big Data to populate a universe of possibilities. Even where the BATNA is not a transaction but is litigation, the artificial negotiator can use knowledge of the law and the outcomes of historical cases to assign the BATNA an expected value. In the field of patent litigation, an algorithm using a multi-factor, data-driven approach has already been shown to be better at predicting outcomes than experienced human litigators.[24]

Just like options, the critical aspect for alternatives comes down to comparison, which means comparing alternatives to each other to find the BATNA, and comparing the BATNA to the options under consideration to judge whether agreement is desirable. To find the BATNA the artificial negotiator can pass the alternatives through the *agreement model* to rank the alternatives for how well they satisfy the client's needs. The algorithm can then reject options that score more poorly in the *agreement model* than the BATNA and focus on options that outperform the BATNA.

---

[23] MNOOKIN, *supra* note 8.

[24] *See* Richard Susskind, Upgrading Justice, Presentation at Harv. L. Sch. at 20:20 (Feb. 14, 2017), ("[Lex Machina] is a system that predicts the outcome of patent disputes more accurately than any human patent lawyer. . . . It knows nothing about patent law . . . [b]ut it's got about 100,000 patent cases, information such as who the judge was, what the courtroom was, who the lawyer was, the law firm, the litigant, the value of the claim, the nature of the claim – a whole bundle of variables. And it transpires that you can make a more accurate statistical prediction of the outcome of the court based on that kind of data than based on legal reasoning, legal problem solving, and other legal methods."), *available at* https://www.youtube.com/watch?v=Vd0PhomzT7g?t=20m20s.

The artificial negotiator can also evaluate all incoming proposals from the other side versus the BATNA score. If despite negotiation the proposals always underperform the BATNA, then the algorithm can recommend ending the negotiation. A similar strategy can be used with regard to the counterparty's BATNA. The negotiator can use available information to brainstorm and identify the other side's BATNA. It can then score it with the *counterparty model* and use it as a benchmark to avoid making offers that are worse than the other side's BATNA and would be unlikely to lead to a negotiated agreement.

5. Commitment

Even before a negotiation is over, commitment is important because the ability to implement an agreement will affect the value of a given option. An offer of one million dollars per week in alimony should meet any divorcee's interests, but is actually a meaningless option because it is infeasible to implement. While all negotiators can remain aware of such gross commitment issues in considering options, an artificial negotiator could take a more systematic approach by calculating implementation risks on all components of a proposed option and discounting the values of such components accordingly. The risks could be estimated based on factors like the historical performance of similar agreements, likelihoods of any legal challenges, and the predicted financial future of the involved parties.

For the aspect of commitment which involves reducing a negotiation to writing, an artificial negotiator may also be superior to humans. First, an algorithm is not going to forget or misremember any of the agreed terms and language. The two sides may still have misunderstandings, but these are more likely to be discovered and resolved than if left solely to human drafters. Second, the artificial negotiator would be able to draft agreements with surpassing levels of detail. Lawyers and others already routinely draft multi-hundred-page settlements and financial agreements in an attempt to cover many possible outcomes. A computational drafter could easily take this to thousands or millions of pages to include almost every eventuality. While such complexity may theoretically lead to more efficient agreements, there are associated costs with adding such complexity. Even if the algorithm(s) can understand it all, human principals could not. A willingness of principals to accept agreements that they cannot, even with extraordinary diligence, fully comprehend would require great trust in artificial intelligence.[25]

6. Communication

So far, I've asserted that artificial negotiators can surpass human negotiators primarily by using raw processing power to consider more interests, more options, more alternatives, and more detailed commitments than humans can handle, and secondarily by using data-driven learning to optimize such considerations in ways that humans cannot. But can an algorithm ever say, "I have people skills?" And does it need to? An algorithmic agent has to communicate somehow with its own principal, which I am assuming for simplicity is a human or an organization with human decision-makers. The artificial negotiator also has to communicate with the other side, which may have either a human or algorithmic agent, but again has a human principal. First, I will consider inter-agent communications, and then principal-agent communications.

If both negotiators are machines, there could be some major efficiency advantages versus negotiations with at least one human negotiator. Computers can share data much faster than people, and can intelligibly communicate much larger data sets. Two artificial systems could exchange

[25] While such trust develops very slowly in the public at large, it is not all that foreign of a concept for experts in certain fields. *See generally* MICHAEL LEWIS, FLASH BOYS: A WALL STREET REVOLT (2014) (exploring how during the 2000s high-frequency trading algorithms were rapidly made responsible for executing over 50% of all equity trades on U.S. securities exchanges). For the public at large, consider the increasing trust that consumers are placing on digital assistants like Amazon Alexa or Google Assistant to make purchasing decisions. *See, e.g.*, Michal S. Gal & Niva Elkin-Koren, *Algorithmic Consumers*, 30 HARV. J.L. & TECH. (forthcoming 2017).

individually "brainstormed" options, for example. This would be a lossless communication, they would simply swap data files, and each algorithm could then evaluate the options according to its own client's *agreement model* and incorporate anything learned from the shared options into its own updated options and alternatives. If the algorithms were permitted by their principals to be fairly transparent, they could even swap *agreement models*, or edited, shareable versions of the models, as a way to communicate interests and priorities that would be more precise than any across-the-table discussion. The artificial agents would need some editing system to avoid sharing certain vulnerabilities and confidential information. These shared models would replace the more crudely estimated *counterparty models*, and except for the withheld information the two algorithms would essentially be processing the same problem. Thus, communication that consisted of merely raw data sharing could lead to natural convergence upon a negotiated agreement (or come close to convergence with a set of narrow ZOPAs) or could identify quickly that no agreement was possible without significant changes in principals' positions.

Principal-agent communication would be largely dependent on the principal's trust. If the principal trusts the artificial negotiator to properly build an *agreement model* and to have sufficient sophistication to appropriately evaluate options and alternatives, then there may be little need for communication beyond the initial disclosure of interests and priorities and early feedback for refinement. As discussed above, disclosure and refinement are critical elements of enabling the artificial negotiator. The human client needs to be able to describe her interests accurately and fully enough for the algorithm to properly construct decision-making code. She would also have to give the artificial negotiator enough information to construct options and alternatives that are reasonable and sufficiently specific enough to her goals. Later refinements could be made through asking the principal questions about the highest rated and most likely agreement options. Thus, the human could reject options that the algorithm erroneously construed would met her interests, or that the algorithm properly constructed but that are not acceptable because of an undisclosed or wrongly prioritized interest. Iterating this process should move the artificial negotiator towards a model that most appropriately aligns with the true interests, priorities, and resources of the client.

A similar iterative process could be repeated once the negotiating across the table is in progress and the algorithm is incorporating information from the counterpart. Additionally, the artificial negotiator could share some of that newly received information with its own principal, so that the principal is better informed about the other party's interests, priorities, and resources. This may lead to a shift in the principal's own positions which could be fed back into the algorithm.

The major problem with all the communication occurring between machine and human is going to be the degree of complexity. If the artificial negotiator's advantage is really in constructing elaborate options and complex models, the principal may be unable, unwilling, or uninterested in understanding all the details. There is a risk that the principal will reject elaborate options simply because of their complexity, which could bias the algorithm towards simpler, non-optimized solutions. Much like a human lawyer, the machine's task will become translation of the complexity into terms the principal can understand and properly consider. A good artificial agent will be able to ask simple questions of the principal despite refining a highly complex model. Ideally, significant refinement will have occurred before the artificial negotiator needs the principal to review options in their elaborate entirety; perhaps this will only be necessary when the parties are close to agreement.

7. Relationship

I have described a communication model that mimics humans in developing and refining understanding, but is ultimately very utilitarian. But what about an artificial negotiator's ability to be empathic with its own principal or to foster trust and understanding with the other side?

Some current computers are already proficient at recognizing human emotions based on visual[26] and audio cues.[27] I would even expect a computer to be better at discerning emotion from written communications,[28] like emails, because of the potential opportunity to compare the text to millions of other written samples pre-classified by emotion. But recognizing emotion is only one part of the task. The machine would also have to decide how to process inputs differently depending on the emotion conveyed along with each input. For example, if a client displays great anger when discussing the sale of an inherited family property, this may indicate that the stated interest in keeping the property is strong and should be highly prioritized. But it might also indicate a more hidden interest in repairing family relationships. It will be hard for an algorithm to resolve such ambiguities. The system may, however, be able to learn that emotions often signal deeper interests and use this as a cue to make further inquiries. An attentive human would do the same.

Another part of empathy is responding to statements and emotions appropriately in order to demonstrate understanding and learn more. Paraphrasing, inquiry, and acknowledgement are possible active listening tools that an artificial negotiator could use. But if the system is just a text or speech interface it may be hard to engender trust with a human principal. An algorithm can select language that conveys empathy and reflects appropriate emotion, but it is less likely than a human to be convincing. The machine will lack emotional legitimacy in the eyes of a human. This could change if humans eventually accept sufficiently advanced machines as capable of sentient thought and understanding. A less advanced machine, but one that is anthropomorphized, may also be able to fool human emotional response simply by having a human look and feel. Additionally,

[26] Raffi Khatchadourian, *We Know How You Feel*, NEW YORKER (Jan. 19, 2015), http://www.newyorker.com/magazine/2015/01/19/know-feel ("[C]omputers can now outperform most people in distinguishing social smiles from those triggered by spontaneous joy, and in differentiating between faked pain and genuine pain.").

[27] *Id.* ("Experts on the voice have trained computers to . . . scan a conversation between a woman and a child and determine if the woman is a mother . . . [and] whether she is angry or frustrated or joyful.").

[28] *Cf. id.* ("Other machines can measure sentiment by assessing the arrangement of our words.").

certain individual principals and certain groups of principals may adopt genuine trust in artificial negotiators and other forms of AI, well before full societal acceptance.[29]

The relationship element also refers to the relationship between the two sides. Once either side is artificially represented it will become more difficult for the sides to express emotion, air grievances, or to feel that they are being heard. As mentioned above, machines lack legitimacy in these empathic tasks. When both sides are artificial negotiators, the negotiation is likely to proceed in a highly utilitarian fashion and without much focus on the interparty relationship. The concept of trust will not be addressed directly, but will be folded into the information sharing that is occurring between both algorithms and will be affected by how each principal responds in the iterative communications with its own algorithmic agent. If a principal severely limits information sharing, the other side will realize that her agent and she are not being cooperative. Similarly, if a principal is specifically encouraging difficult tactics, the opposing side will recognize an unusual pattern that doesn't trend towards both algorithms having models with some overlap.

Perhaps the strangest scenario in terms of interparty relationship is the case when a human negotiator is dealing with an artificial one. At least for face-to-face negotiations human negotiators trade, in part, on their personality and their ability to understand and influence the other side. But a human may not be able to influence a machine in ways that don't involve explicit disclosure of information. For example, one common skill used in negotiation is to reframe a certain position to get the other side to recognize its importance and value. The listening party may soften towards the position as it is repackaged and reemphasized. Contrastingly, a machine listener hearing a similar point three or four times may not respond to the follow-ups as new information at all. The

[29] *See supra* note 25.

repetition itself could be taken as a cue, but very sophisticated language processing might be needed to categorize each reframe as a unique piece of new information.

Conclusions

I have described an artificial negotiator which can carry out complex negotiations on behalf of a client, and exhibits some proficiency in navigating each of the seven elements of negotiation: interests, options, legitimacy, alternatives, commitments, communication, and relationship. The artificial negotiator is particularly adept in areas where its analytic power can be used to consider large datasets, such as brainstorming and evaluating detailed options and alternatives, aggregating and organizing objective criteria, and drafting detailed commitment terms. The artificial negotiator's shortcomings surface in areas that, perhaps unsurprisingly, highlight its artificiality. This is most notable in the tasks of developing relationships and trust with human clients and counterparts. For communication, the artificial negotiator could have some efficiency advantages speaking with other machines, but would struggle to translate internal complexities into human terms.

The core of the artificial negotiator I have described is its function in learning both sides interests and translating them into decision-making models with which to judge options, alternatives, legitimacy, and commitments. The interests themselves are the core of this model, but the artificial negotiator must rely on its ability to communicate and have a well-functioning relationship with humans. I believe that technological advances will provide the computing abilities for artificial negotiator to excel in complex and varied negotiations. However, to truly surpass humans, artificial negotiators will need not only to possess these processing powers but to develop greater empathic capacities, so that they can become genuinely trusted by the principals they will serve.