THREE LAWS OF AI TO CONSIDER

During a recent panel discussion on artificial intelligence (AI) and the law, the issues of trust and responsibility were considered. One of the questions from the audience asked what three laws of AI might be, a la Isaac Asimov's The Three Laws of Robotics within his Robot series (and, eventually, the Foundation series) of science fiction books. While perhaps not as earth- or empire-shaking as envisioned by the late great writer (and hence the lower cases), considering a potential three laws of artificial intelligence, I found myself reflecting upon past lessons learned from our last several decades. And it seems to me that we can be—and should be—careful, considerate, and successful in the midst of what some are describing as a new frontier, which it very well may be, but we shall see.

We really don't need to abandon all that is present and erroneously seen as old to charge forward with the new. Backward compatibility is an obvious must-have that is all too easy to overlook. Why is it obvious? Well, in the context of people's current computers, for instance, it is because the vast majority of people have older machines and they quite arguably don't need the newer capabilities, especially ones that hinder the ability of existing ones to operate. Is it possible that they will grow into the need for these capabilities if those capabilities are introduced in a thoughtful manner? Sure, of course. But that thoughtfulness does not include ignoring the fact that they have perfectly good machines in the here and now. If we are open and thoughtful from the outset and do the groundwork early, then we will have a framework for complete success. That will save a lot of downstream costs and anxiety. And isn't that what we are working to do? Shouldn't we be working on behalf of people to improve their lives rather than trying to squeeze them while making their lives worse? That is clearly a question that needs to be answered affirmatively, but if we are not even getting to this question, be it due to obfuscation or simple neglect, then it becomes easy to overlook.

Related basic rules of artificial intelligence have general applicability as well. Yes, just because we can do a thing does not mean that we should do a thing or that we can do these things better. If we had thought in advance as a society, for example, about the outsourcing of our jobs some several decades ago, and how to transition our workforce in a healthier manner, then we would have avoided a great

deal of personal and social upheaval. Displaced workers end up being greater burdens on us when their outcomes are deeply negative. We as a society end up bearing the impact of the despondency of those left behind. The people pushing the benefits, essentially to themselves at the so-called top, are the people who are also the best-suited to avoid the repercussions, for better and for worse. If they help us do so for the better, with a bit more thought and planning, we will create even more abundant and widespread opportunities. Yet, in a less thoughtful environment, those of conscience and ability willing to expose themselves to these concerns are the ones who primarily end up getting scapegoated. And to what end? Are we setting social policy to promote the most vicious, the most cannibalistic, the most profiteering among us? Or are we trying to do other things? Those are some starting thoughts.

Here are my three laws of AI. The first law might sound somewhat odd. It's "Maybe this can be useful." This law is a mindset. Be careful as you barrel forward. Don't, to utilize my first example above, overload our perfectly good machines that are 100% doing the jobs that we are asking them to do, and doing those jobs well, with bloatware. We have seen bloatware used to subsidize more affordable computers, which is fine if a bit annoying, but now to see it employed in mass to brick nearly everything is, in a sense, not just unconscionable but quite wasteful. No, this is not the direction we need to be moving in, or at least this is not the way we should be doing it. Subsidize the advancements without victimizing the vast majority and creating colossal amounts of waste. This can be done. The first law and the mindset it promotes foster basic considerateness. We must have sufficient, if not ample, amounts of humility not just to operate within a society, but to interoperate successfully and for everyone's benefit. This can unequivocally be done. Mindsets are quite powerful things.

Second, "If useful, this is going to take a lot of work to do properly." Sure, this is another mindset, but at this starting point, we should be examining approaches more than focusing only on monetizing specifics, as those profit centers will certainly come, and they will arrive in sustained force if we do this right. If large-scale data analysis can give us benefits, then take the process stepwise. Ensure that people know what you are doing. Open up your processes to teaching rather than creating closed

ecosystems that protect monopolies on knowledge that has no reason to be monopolized other than fearful desires for control and hoarded benefits. This also is obvious, but it needs to be articulated. Bring people with you on this journey to a better place, and that place will be sustainable.

And remember, just because it seems faster and newer, that doesn't mean we should or need to abandon people or ways. Think about how worthy local farm-to-table cooking is. Yes, modern agriculture is a great thing, but then, there are also great reasons to take our time and be deliberate about how we go about doing the most basic of things. Capabilities will change and improve, but value throughout will persist, even if in our haste and neglect we unnecessarily trample these things. You achieve multifaceted success through a bit more patience and work, but in the end you lose less and gain more. Let us not cause the loss of knowledge and gratuitous pain only to find in the end that we could have attained the same benefits but for some acknowledgment of thought.

This entails a more thorough consideration of the landscape of people than what we have seen in the past decades with waves of outsourcing, unsustainable financial practices, and simply general tunnel-vision assumptions that one's own perspective is prevalent and even singular. That is simply not the reality of how things work. If we make these mistakes now and forget that we need to put in the work, then we will pay for those mistakes downstream. And the game does not have to be about pushing off the costs of those mistakes to other people. Instead, if the promise of what we are looking at is real, then we can and need to do this properly.

We can involve people in developing the technology that is meant to make their lives better. Isn't that, after all, definitionally, what we must do to achieve the promise of a technology that can in innumerable ways improve our lives? If we can't even prevent the harmful and destructive imposition of that technology that purportedly improves our opportunities, then what is it that we are doing? The theoretical envisioning of what AI can do is supposedly pervasive. As such, we need to bring everyone along, because everyone has a stake in the success, and is a part of defining that success. Jobs will be created, and we need to do the work of defining those jobs. We need good teachers

who understand what it is that we are doing and trying to achieve. We need people to connect with those jobs that will in no small part be about understanding our current processes and systems, and improving and adapting those processes and systems, and the people who deliver them, over many years. This is an inflection point requiring deliberation.

This is not some kind of fire sale. We are looking at generations of steady opportunities and what is in front of us right now, to extend again the original case in point, is not an opportunity to load our existing machines with bloatware so that we can prematurely obsolete equipment and attack wallets like much of the arm-pulling slot-machine crowd-sourced profit-mongering behavior that has pervaded so very many of our systems such that the vast majority of us in our hearts are filled with deep disdain and, ultimately, disengagement. Put those emerging capabilities in the cloud so that existing machines can keep doing their effective work, and think about the people we are serving. Teach. Engage. That is the way to do this properly. It will take time, and it will be worth the effort.

Third is, in essence, a reiteration of the first two laws, or at the very least a natural extension of the second one, which, if done correctly, should make this third one readily accessible: “Remind people of where we came from.” We should do this because there is an ultimate power in understanding and in context, and it’s the freest thing in the world to deliver to people. We must not lose sight of the feeling of openness that interconnectivity provided to us and keeps providing. We are not looking at sea-change obsolescence. That kind of mentality is just wrong. Maybe some limited improvements can be made, but the truth of the matter is that when we forget how we got to where we are, then we are operating without context or direction. And that is how we have elevated people with the trappings of finery and high expectations, but we are also left wondering why certain problems seem to be increasing.

Impose AI haphazardly, and those exact problems will keep amplifying. And those compensated to obfuscate, in a heavy paradox or our modern time and system, when we seek transparency and an open discussion, will work at odds with improvement in the very names of the liberty and openness that we seek to achieve on our way to increasing and universalizing happiness. But isn’t it strange

that with all these capabilities, people seem now more discontented than ever? That those suffering the most suffer now even more? We need to remember where we're coming from so that we have a direction to aim for. This is especially true with AI, because aiming massive computational and data-retrieving capabilities haphazardly will end up, in the immediate and most obviously, divesting people of their livelihoods and creating large amounts of waste when we leave behind people in the direction of forcing greater computational and data-retrieving speeds as a singular metric of consideration.

Now, another more important metric is, what are people doing right now, and how can they do these things better in the sense of making them happier and giving them stronger understandings of what they're doing and how it fits in the broader landscape, and in the sense of helping the rest of us to understand what they're doing and creating a much more robust society that is interwoven through knowledge, action, and self-starting inspired synergies? Those capabilities are considerably more important than trying to push big chips in mass, and if you're really listening to the leaders of the hardware of this technology, they completely understand these realities. Don't let the pencil-pushing bean-counting get in the way and create mission creep. Yes, we can absolutely sustain the emerging technologies and growth, but we don't need to do it on the backs of everyone, or at least the overwhelming majority of people that we need to bring with us to make the endeavor successful.

Those are my thoughts on what three laws of artificial intelligence might look like in the here and now in our latest brave new world. I suspect that the new stuff really does kind of look like the old stuff, though. In my forays through the private and public spheres, I would say that the most important parts are not the maneuvering, or the moving fast and breaking things, or even the process of repair, but in the consideration of how we conduct the journey. And I think, yes, maybe this stuff will be useful, we will need to do a lot of work to get there well and properly, and we should be constantly doing look-backs and assessments to help us define and optimize the best way forward.

BIO

Darcy Paul

Attorney, Paul Law Group, A Professional Corporation

Darcy is a graduate of Harvard College (B.A., biology) and Harvard Law School (J.D.), where he served as editor-in-chief of publications covering the areas of scientific research and legal issues related to technology. He was twice elected to the Cupertino City Council, where he served from 2014 to 2022, and served as Mayor of Cupertino in 2018, 2021 and 2022. An attorney for more than 20 years, Darcy is a patent attorney, but a generalist. His law practice is located in Silicon Valley.

ABOUT CONSTELLATION RESEARCH

Constellation Research is an award-winning, Silicon Valley–based research and advisory firm that helps organizations navigate the challenges of digital disruption through business model transformation and the judicious application of disruptive technologies. Unlike the legacy analyst firms, Constellation Research is disrupting how research is accessed, what topics are covered, and how clients can partner with a research firm to achieve success. Over 350 clients have joined from an ecosystem of buyers, partners, solution providers, C-suite, boards of directors, and vendor clients. Our mission is to identify, validate, and share insights with our clients.

Organizational Highlights

- Named Institute of Industry Analyst Relations (IIAR) New Analyst Firm of the Year in 2011 and #1 Independent Analyst Firm for 2014 and 2015.
- Experienced research team with an average of 25 years of practitioner, management, and industry experience.
- Organizers of the Constellation Connected Enterprise—an innovation summit and best practices knowledge-sharing retreat for business leaders.
- Founders of Constellation Executive Network, a membership organization for digital leaders seeking to learn from market leaders and fast followers.

www.ConstellationR.com

@ConstellationR

info@ConstellationR.com

sales@ConstellationR.com

San Francisco Bay Area | Boston | Colorado Springs | Denver | Ft. Lauderdale | New York Metro
Northern Virginia | Los Angeles | Portland | Pune | San Diego | Sydney | Washington, D.C.