



Reflections on AI

Q&A with

Prof. Patrick Lin

“AI is opening up the space frontier...and this seems to be a clean slate ... a chance to remake human society in the cosmos with ethical forethought.”

The [TUM IEAI](#) had the pleasure of speaking with Patrick Lin, Director of the Ethics and Emerging Sciences Group, based at California Polytechnic State University.

Q: What is the biggest misconception about AI?

A: In my mind, the biggest misconception is perhaps that AI “thinks” or “understands” or has “intent”. Maybe it’s natural for us humans to anthropomorphize things—we seem to be hardwired to do that, e.g., to see faces that aren’t really there; this was an evolutionary advantage. But the tendency is leading to hysteria and overblown fears now, such as the possibility of “superintelligence” or associated existential risks of AI taking over the world.

To be sure, there may be plausible existential risks, but probably not the sort that has been more commonly posited, e.g., the bizarre “paperclip maximizer” scenario. For instance, if AI were put in charge of nuclear weapons, then it may do something unintended and existentially dangerous, but that’s more of a mistake of humans to put AI in charge of such systems in the first place.

Q: What is the most important question in AI ethics right now?

A: One of the most important questions in AI ethics is still the issue of overtrust. As suggested above, overtrusting AI is easy because of our own anthropomorphic mistakes, that is, when we attribute human-like capabilities and intentions to AI when there really is none.

Overtrust is also easy to do when an AI system seems to be working reasonably well, even though a great intransparency remains into how or why AI does what it does. To many, even technical researchers, it’s basically magic. But without understanding the factors AI considers or uses in its reasoning, and to what degree, we’ll keep seeing serious errors, such as biased decisions that affect real human lives. Some of these bad decisions can be fatal, such as misclassifying a person in a way that makes them into a target.

One example of overtrust is to prematurely deploy self-driving systems without sufficiently understanding the edge cases, including how and why errors may occur. A more serious example is to field security and defense AI prematurely or without thinking through the guardrails—that’s even more irresponsible and dangerous given the stakes in armed conflict and the very real possibility of escalation.

For instance, think about large language models (LLMs) right now [for use in intelligence gathering]: they still “hallucinate” or make stuff up, and it’s easy to “poison the well” by slipping in bad information into its training dataset, in this case, the open web. This intelligence failure could cause a false alarm and divert precious resources away from other priorities, or it could cloak the plans of a real attack, causing us to be unprepared for maximum damage.

Q: What are key ethical responsibilities in terms of exploration of and policymaking for outer space?

A: AI is one of many enabling technologies in humanity’s push into outer space, just as it is for many other domains. But, as they say, with



Reflections on AI

Q&A with

Prof. Patrick Lin

“AI is opening up the space frontier...and this seems to be a clean slate ... a chance to remake human society in the cosmos with ethical forethought.”

great power comes great responsibility. As more nations have or are developing launch capabilities, outer space is becoming more crowded. Because it's more congested—whether it's about Low Earth Orbit (LEO) or competition to race toward an asteroid or particular spot on the Moon—it's also becoming more contested, given the still tense geopolitical environment on Earth.

This means we need to anticipate new conflicts, so that we can identify and prepare “off-ramps” to deescalate conflicts before they go too far. The stakes are very high: besides global positioning systems we use every day, we also rely on space capabilities and data for, say, weather forecasts, Earth observations (for everything from better managing farmland to monitoring natural disasters and military movements), as well as precise timing for financial transactions, mobile, and computer communications. With a space conflict, all that can be affected or even destroyed—everyone would lose.

These critical systems don't even need to be intended targets in a space conflict; they can be collateral damage as conflicts accelerate the space debris problem. Therefore, I think one of the most important questions facing us in outer space conflict is this:

When is our responsibility to not create more space debris greater than our responsibility to protect? That is, even if a state has a right to self-defense in outer space, can they also have a duty to not exercise that right because the stakes of escalation might be so high? Every armed conflict has a similar problem—is retaliating a proportionate response?—but not to this degree in outer space. It's possible that we may need to let some serious violations or even attacks go unanswered in outer space,

which we might not otherwise on Earth, so to not imprison the Earth in a cocoon of debris—which is a minefield for launches from Earth, as well as for space assets already in orbit—and this is a very unfamiliar scenario in human history.

Q: How is technology (and AI in particular) changing the current landscape and ethical dimensions of activities in outer space?

A: For outer space, AI is allowing us to better maneuver around while in space; this greater control, helps us in our explorations and other activities—both for good and bad. Reusable rocket boosters are a huge cost-savings and are enabled only by AI for precision calculations and control at high speeds, e.g., to land a falling booster safely on a small pad. Other space vehicles, such as rovers, benefit from AI in similar ways as Earth-bound vehicles, such as autonomous control.

With AI and other technologies, deep space missions are now more feasible. For science more generally, AI can help us identify and characterize asteroids, moons, planets, constellations, black holes, space signals, and so on more precisely, some of which was previously impossible without AI. All of this, by the way, can also have economic and other benefits.

In terms of ethics, AI is opening up the space frontier—which has been called the “final frontier”—and this seems to be a clean slate. It's a chance to remake human society in the cosmos with ethical forethought this time, drawing from the terrible lessons we've learned at great cost throughout human history.



Reflections on AI

Q&A with

Prof. Patrick Lin

“AI is opening up the space frontier...and this seems to be a clean slate ... a chance to remake human society in the cosmos with ethical forethought.”

Q: What is the role of academia when it comes to the regulation of trustworthy and responsible AI?

A: AI is complex, but it is not magic. Academics, both technical and non-technical, can help to explain how it works, what it can do for us, and what risks it might pose in simple, understandable language. That’s what academics do in teaching: we take complex ideas, and we translate them as education for the next generation. On the research side, we can go much deeper into the analysis than industry and government can or even wants to do, and this is critical in drawing out important nuances that might be missed on a superficial discussion.

Also, academics can better resist industry and governmental pressures, even though conflicts of interest are still possible. I don’t know if anyone is ever really free from bias, but perhaps academics are best positioned to be neutral or objective given a certain distance from the corrupting influence of money. Academics can also be free to be activists, or to be the intellectual lead for such activism, where they strongly believe there is an injustice that needs to be remedied.

Finally, remember that academia is largely where the call for ethics started in the first place! This includes trust, responsibility, and other related buzzwords. The ethics siren certainly didn’t come from industry, even if some companies today make a show about standing in front of legislators, practically begging for regulation, while behind our backs, they race to commercialize and deploy AI systems that no one really understands, including themselves when they don’t have the benefit of an open, public conversation and academic investigations.

That said, I know there are some ethicists embedded within industry who are trying to help steer those ships in a responsible direction. We need to work the problem from both directions: from the outside as well as from the inside.

Q: We often say that AI is changing the world. To what extent is AI changing us as humans?

A: I’m not really sure, but one thing I might note is that there seems to be a long-standing interplay—a feedback loop—between humans and technology. We simply would not exist if it weren’t for technology, specifically in making and using tools to manipulate the world to better suit our frailties. We don’t have the natural capabilities and advantages that other animals have, and the only thing keeping us alive is our ability to use and make technology, along with our sheer tenacity.

Where, say, simple cooking for survival has evolved and transformed human life into many different cuisines, traditions, and cultures; or where simple clothing for survival has evolved into entire industries, sweatshops/other exploitation, fashion, and culture, I think we can expect important technologies to transform our world.

For those of us old enough to remember the dawn of the computing age: from a satellite view, ok, maybe human life on Earth looks more or less the same; but on a closer view of our relationships, how we work, how we make war, and so on, the difference is revolutionary. AI seems to be poised to have a similar kind of impact. But no one really knows—we’ll just have to find out when we get to the future...

Reflections on AI

Q&A with

Prof. Patrick Lin

“AI is opening up the space frontier...and this seems to be a clean slate ... a chance to remake human society in the cosmos with ethical forethought.”

Meet the Expert:



[Patrick Lin](#), PhD, is the director of the Ethics + Emerging Sciences Group, based at California Polytechnic State University, San Luis Obispo, where he is a philosophy professor. He also currently serves on the US National Space Council’s Users Advisory Group and is affiliated with Stanford Law School, Czech Academy of Sciences, World Economic Forum, Aurelia Institute, and other leading orgs. Prof. Lin is well published in technology ethics—including on frontier development (esp. outer space and the Arctic), AI, robotics, cybersecurity, bioengineering, nanotechnology, security technologies, and more—and is regularly invited to provide briefings on the subject to industry, media, and government.

Disclaimer: Please note that the text version of this interview has undergone slight edits for clarity and conciseness.