EconoFact Chats: Managing AI Risks
Josephine Wolff, The Fletcher School, Tufts University
Published on 5th April 2026

Michael Klein
I'm Michael Klein, executive editor of EconoFact, a non-partisan web-based publication of The Fletcher School at Tufts University. At EconoFact, we bring key facts and incisive analysis to the national debate on economic and social policies, publishing work from leading economists across the country. You can learn more about us and see our work at www.econofact.org.

Michael Klein
Artificial Intelligence is seen as a disrupter in the best sense of the word. It promises to provide new capabilities that will replace less efficient methods of obtaining and acting on information for a wide range of commercial and personal uses. But AI has the potential to disrupt in more damaging ways as well. Without human input, the actions driven by AI could lead to bad, or even disastrous outcomes. Companies are increasingly seeing AI as an essential tool for their operations – but they are also aware of potential problems and are trying to find ways to insulate themselves from the downside risks of using AI. To discuss these issues, I am very pleased to have as my guest on EconoFact Chats Professor Josephine Wolff. Josephine is my colleague at the Fletcher School at Tufts University where she is a Professor of Cybersecurity Policy. She has published two books with MIT Press. The 2018 book with the striking title *You'll See This Message When It Is Too Late: The Legal and Economic Aftermath of Cybersecurity Breaches* and the 2022 book *Cyberinsurance Policy: Rethinking Risk in an Age of Ransomware, Computer Fraud, Data Breaches, and Cyberattacks*. Josephine, welcome back to EconoFact Chats.

Josephine Wolff
Thank you so much for having me, Michael.

Michael Klein
So let's start out with a basic question to set the stage – what is artificial intelligence and what are some of its current and prospective uses?

Josephine Wolff
So when we talk about artificial intelligence broadly, we're usually talking about some version of building computer systems that imitate the way that humans think and act. And that can be everything from robotics – to build robots that can dance or manufacture or do other things that we associate with humans, or more, more commonly, we think of it as being some version of machine learning. Where you take a whole bunch of data points or inputs about people, about weather, about car crashes, about anything that you're sort of interested in making a prediction or decision about, and you use this machine learning set of algorithms of math essentially to find all

possible patterns across those data points and make those predictions about how many car crashes will there be next year? Or how long will somebody live? Or what will the weather look like next week? Or should this person be released on bail, what's the risk of recidivism if they are? And so we're seeing these algorithms being used to assist or in some cases, even make decisions across a whole range of different domains. What should we price your car insurance at? How should we make decisions in the criminal justice system based on the patterns that these machine learning algorithms can extract from data sets?

Michael Klein
So there are advantages for these uses, but there are risks as well. Let's start with one example – self-driving cars and trucks. The benefits are reduced costs from not having to pay drivers, and they have to be weighed against the risks – the costs of the possibility of more accidents. What do we know about the benefits and costs of this technology as it exists today?

Josephine Wolff
So it's a great question, and I think it's a challenging question to answer very concretely, because the technology itself is still evolving. So when we think about autonomous vehicles exactly, as you say, there are these risks that we don't exactly know how to calculate, because we haven't had autonomous vehicles very long, they're only on the roads in a few places. We think, based on the data we have so far, that they're a lot safer than human drivers. They don't get tired, they don't get drunk, they don't, you know, make the kinds of mistakes or panic decisions that human drivers sometimes make. On the other hand, there's this sort of unquantifiable piece, which is, what if somebody compromises the autonomous vehicles? What if, instead of the sort of random bad decisions and mistakes made by human drivers, there's a concerted deliberate effort to crash a bunch of cars, or even just crash some specific cars with certain people in them. And so I think the trade off here is the idea that we could save a lot of lives, we could make driving much safer and much easier and perhaps even much cheaper, as you say, for a lot of people. But there's this very difficult to measure what if scenario, of what if the autonomous vehicles either malfunction in some way we don't understand or don't predict, or are compromised remotely, because they're going to be so connected to computing power in order to do the processing they need to do, that they're able to be compromised and kind of crashed in the sense that we're giving up our control over it. Instead of, you know, me relying on myself and my own good judgment and my driving, it's now going to be a little bit more at the mercy of a technology that's very hard to understand.

Michael Klein
Well, Josephine trucking companies have liability insurance to protect themselves against claims from accidents that may or may not be due to driver's errors. Could that same type of insurance be used if and when these companies rely on self driving trucks?

Josephine Wolff
So that's a huge question right now in the insurance industry, because artificial intelligence is going to touch pretty much every type of risk that is currently insured. And so the insurers right now are grappling with this question of first of all, should we be excluding some of these AI risks from our existing coverage? Because all of their actuarial models are built around cars that aren't autonomous…are built around systems that don't incorporate AI to a great extent. And what we've seen so far is that most insurers have not excluded artificial intelligence from their commercial general liability policies, or most of their other policies…[a] couple have, but that doesn't seem to be the trend so far. On the other hand, they also haven't suffered very many losses or claims related to AI yet. So I think it's still a little bit of an open question. As these risks sort of accelerate, as we see the AI adoption continue to grow, will insurers be willing to cover this? And if not, are they going to start offering separate AI insurance, which is what we saw in the cyber insurance domain? Or is there going to be this gap in people's insurance coverage? And if so, what are they going to do about it?

Michael Klein
I read a recent article that you co-wrote with Daniel Schwartz, a law professor at the University of Minnesota. I found it really interesting, something I hadn't thought of before. You discuss how tort law and liability insurance can, in some instances, serve the same role as government regulation, and in fact, in some ways, this private sector solution may be preferable from a public policy perspective. Can you explain why this would be the case?

Josephine Wolff
Sure, so I think this is often something that we hear in the tech policy space, partly because there's this idea, and I think a fair idea, that technology moves very quickly, and lawmakers and policymaking moves very slowly. And so when you think about how do we address these risks from emerging technologies, often people in industry, and people who work on the tech side argue that policy is not going to be able to catch up. And instead, especially in the United States, where we are a rather litigious society, there's this idea that what you want to do is, when something goes wrong, when there's an accident or a security or safety problem, you should sue, and the courts, through tort liability, will help assign who's responsible for this, and that, in turn, will place a lot of pressure, or create some incentives for the companies that are making these technologies to make them safer and figure out what needs to be done to avoid that liability in the future. And part of the idea is that this will put the safety measures in the hands of the companies and the technologists who understand the technology the best, instead of relying on government to come in and say, here's how you should do this. And it will also make it easier for them to update those safety practices and measures. If the technology changes in the next six months, the next year, then you're not waiting for Congress to pass a new law or for some sort of new lengthy regulatory rulemaking process. You're instead able to just kind of adapt and change

your practices as you go. So that's the rationale behind imagining that liability can help shift technology in a safer direction without having to wait on the somewhat unwieldy policy process.

Michael Klein
And insurance companies have an important intermediary role in this as well, right?

Josephine Wolff
Absolutely. Right, so when you think about how that process is going to happen first, anytime there's any kind of expensive risk, companies are going to go to their insurers and say we want coverage for this type of risk, and the insurance companies are then going to have to decide, first of all, are we willing to cover this? And if so, what kinds of measures are we going to put in place to protect against moral hazard? How are we going to make sure that if we give you coverage for this risk, you don't run straight towards it as fast as you can and just embrace it, because now you're not going to have to pay when your autonomous vehicles crash or something like that. And so in order to do that, insurance companies do two things. First, they have sometimes deductibles, but the other thing they do is they usually try to collect some data on what are the ways that we can make this technology safer, and then they require that of their policyholders. So right, we see this in cars where you have insurance companies back in the day collecting data about seat belts, about air bags. You see this around things like fire insurance, where there are requirements to have sprinkler systems or smoke detectors or things like that in order to be eligible for fire insurance coverage. And that would be the hope with other emerging technologies, that insurers would be able to say, okay, here's the equivalent of seat belts for AI or smoke detectors for AI, and require that, so that you would actually have insurers kind of helping disseminate these safeguards and safety measures through their coverage, through their policies.

Michael Klein
And in fact, in this case, it could be that it's company specific then, rather than sector wide, so it could really be closely tailored to the risks that certain companies face, right?

Josephine Wolff
Exactly, and also, insurance policies typically renew every year, right? So back to the idea of kind of updating this regularly. The idea would be that the insurance companies are in a good position every year to say, okay, the technology has changed, here are the new requirements, here are the new ways that we're going to deal with this moving forward.

Michael Klein
As I said, I found that really interesting, and not thought of sort of that role of tort and insurance companies in that way, but it was a really interesting way to think about it. So in that article, you argue that certain types of AI risk, like autonomous vehicles or failures of AI enabled medical devices might be relatively well suited to having this kind of liability and insurance approach to

promote the safe use of AI. How would this work in the case of autonomous vehicles, for example?

Josephine Wolff
So I think it would ideally work pretty closely in parallel to how it worked for regular cars, which is that you have a fairly high volume of car related incidents, because there are a lot of cars out there. And every time there's any kind of incident, pretty much it gets reported to an insurer. So we don't have the issue, which we have a lot in cybersecurity and some other areas of AI, where people are keeping incidents secret. They don't want anyone to know that there's been a data breach, that there's been some kind of AI malfunction, because they think it could be bad for their business. Car accidents, pretty much always you get a report…you get some kind of insurance claim, so you can collect a lot of data on where are autonomous vehicle incidents happening? Why are they happening? Is it an issue with the computer vision that they can't see certain things at night? Is it an issue with certain parts of the roads where the lanes are less well marked and it's harder for the computers to identify them? Is it related to certain types of obstructions? Is it related to certain types of software? And the idea would be that with that volume of claims data, the insurers could then do their own analysis to figure out, okay, here are the safeguards we need to have in place to try and make these types of accidents less likely, right? And again, it's analogous to what the insurers helped do around seat belts, around air bags, around other types of vehicle safety measures. I think that seems very doable for autonomous vehicles, because there's going to be a lot of data. It's going to be available, and I think it's not unrealistic to expect that insurers could play a large role in collecting and analyzing those data sets and then sending back out into the industry some instructions about, okay, you need to be implementing these types of safeguards, or you need to be using this type of software, or running these types of tests on your software in order for us to feel comfortable covering you. And because also, auto insurance is something that every single person on the road has to buy, I think it's going to be a very effective dissemination tool, because you're not going to buy an autonomous vehicle if an insurer is not willing to cover it, because our entire sense of sort of how we are paying for car crashes is based around insurance.

Michael Klein
You also argue, in that article, that other risks arising from the use of AI don’t lend themselves to being minimized through this approach. For example, you talk about the spread of misinformation driven through the use of AI. Can you offer an example of this type of risk and who would be adversely affected, and why it's not lending itself to this tort and insurance approach?

Josephine Wolff
Sure, so I think part of our point here is AI is going to be used, and is already used in a huge number of different ways. And the risks that we were most worried about in terms of insurance

and liability not being able to help with risk mitigation in a meaningful way for AI are those where the adverse effects are very hard to quantify and measure. I think disinformation is a good example of that, where if we're seeing a huge influx in AI generated disinformation, we might have a strong sense that that is harmful in variety of ways, undermining trust in institutions, perhaps influencing elections, perhaps making people believe very incorrect and damaging things about health for various alternative medicines or anything else, but that's going to be hard from an insurance standpoint, or from a tort liability standpoint, to really quantify how many people have been directly affected, how effected they've been, how much of the disinformation that they've encountered is AI generated. Right, another thing that's very different here is in the autonomous vehicle example, we know when a car that gets into an accident is being controlled by artificial intelligence. That's an easy to discern thing. We don't necessarily know when we're looking at information, text posted online, whether or not it's been generated by artificial intelligence. And so if you're trying to separate out which of these risks are associated with AI and which aren't, that's going to be more challenging. And we don't have the same kind of ability to measure safeguards, right? So if I said to you, you know, we're gonna take every example of disinformation and put it into a data set and try to analyze which of them, you know worked well, which of them didn't…what we can do to protect people better in the future…that's much harder to do than it is with the autonomous vehicle example, because there aren't the same kind of parallels around you know, okay, was the software tested to look for bicyclists at night? Was the software tested to look for small dogs that weren't crossing at a crosswalk? You know, all of the different ways that you might try to enumerate the potential risks and the ways that the software could be tested or safeguarded against them, I think will be much harder when we're talking about things like disinformation, when we're talking about a whole variety of AI risks, around catastrophic risk, around malfunctions that are fairly low frequency, and so we can't build huge data sets of them around issues where it's hard to identify how involved AI was, right? If you think about cyber attacks that have used malware generated by AI, again, we're not necessarily going to know when we see malware, how much was AI used to write this? To what extent is this something that we should sort of separate out and categorize under AI risks. And that's going to make it a lot harder for insurers to build these models and to do the same kind of risk mitigation.

Michael Klein

So one issue here is the risk could be widely shared, and then you would need to have a class action lawsuit. And we do have a current example of a class action suit because of the use of AI. Anthropic has to pay for a copyright infringement of a billion dollars to a very large number of authors, myself included, maybe you also, whose work they use without obtaining permission. Can you explain this case and whether it's likely to result in a more responsible use of AI as a liability insurance model might suggest?

Josephine Wolff
Yeah so I think it's a really interesting case. It's also, I would say, you know, a case that insurers have been tracking very closely, because they are on the hook for covering a lot of these legal costs. And it's a case that really gets at the heart of training these large language models. Because Anthropic, like most companies, models are trained on basically all the text that's available on the internet, which is most published works, and a whole bunch of other things that have been protected by copyright. And so it's a class action lawsuit in which a number of authors and publishers have said, hey, you can't just take our material, use it to train your model and not license it appropriately, right? You can't just decide because your AI copyright doesn't apply to you. And it's, I think, a huge risk, not just for Anthropic but for everybody out there who's building large language models, because we know that these models require an enormous amount of text to be trained, right? You can't just say, okay, I won't use any of the copyrighted text. I'll just use this small basket of unprotected text that I'm able to find somewhere else. And so it's really at the heart of can we have large language models? Will it be possible to continue to build them? And I think that's the core reason why Anthropic decided to settle, because a ruling against them would have, I think, cut deeply at the very core of what they do. But there is, there is a huge risk still here for other companies, because the settlement doesn't actually resolve the underlying legal issue.

Michael Klein
There's another lawsuit with Anthropic that is currently going through the courts, and it has to do with Anthropic and the Department of Defense. Can you describe what's going on there please?

Josephine Wolff
Yeah, so Anthropic had reached an agreement with the Department of Defense to provide AI tools to them for intelligence analysis and other purposes, and then they got into a dispute, and the contract negotiations fell apart. Ostensibly, Anthropic says because the government wanted to be able to use their tools for 'all lawful purposes,' and Anthropic wanted reassurances that the tools wouldn't be used for broad domestic surveillance, and wouldn't be used for fully autonomous weapons, that is, autonomous weapons that could be launched without a human signing off on the decision. And the Department of War wouldn't agree to that, and so instead, they went and they negotiated a contract with Open AI, but they also…the government designated Anthropic as a supply chain risk, which meant not only that they weren't going to use Anthropic tools, but that essentially, no government agency, no contractor that wanted to do business with the government could use Anthropic, and so Anthropic sued because they said that designation, which is a designation that in the past we've only ever seen applied to foreign companies that we think might be undermining the national security of the country in a really profound way, was retaliatory…was the US government sort of trying to hurt them.

Michael Klein
So one aspect of this points to the nightmare scenario that AI could lead to some kind of catastrophic risk like triggering a nuclear conflict, or a global pandemic. And there are reasons why catastrophe insurance doesn't work. Insurance is based on idiosyncratic risk. Not everybody gets in a car accident on the same day, but a flood or an earthquake affects a very large number of policyholders all at once. First of all, Josephine, how likely is it in your opinion, that there would be some type of AI caused catastrophic event?

Josephine Wolff
So I tend towards optimism on this front. I think it's, of course, always possible, but I don't see a lot of science to suggest that AI catastrophe is imminent.

Michael Klein
Okay, that's reassuring. So even if you're optimistic, you might want to see some regulations in place to minimize these kinds of events. And is it reasonable to think that regulation could, in fact, play this important role?

Josephine Wolff
I think it's reasonable to think that regulation can certainly help, right? And I think that things like transparency requirements for reporting incidents and issues around AI, and various types of testing and audit requirements to make sure that these systems are being really rigorously assessed before they're released out into the wild can certainly help bring down the risks associated with them.

Michael Klein
So Josephine, I mentioned in the introduction that you're a professor of cybersecurity, and you've published a couple of books on the topic. Can you discuss the parallels, or the lack of parallels, between the risks from cyber attacks and from AI, and how policies meant to minimize these two types of risks could be similar, or would necessarily differ.

Josephine Wolff
So I think there are a lot of parallels, right? When we think about the emergence of computing technology and the internet, there are a huge number of risks associated with that. It permeates every sector of the economy. Everybody starts using computers. Everybody starts using the internet. And there are a bunch of risks that we don't really anticipate, and those range from fraud to ransomware and extortion, to denial of service attacks to data breaches. And as time goes by and we see more of these incidents and we're able to collect a little bit more data about them, we start being able to sell insurance. We start being able to resolve some of the liability claims. We start being able to make better recommendations about how should we safeguard these systems. I think…I hope that AI will follow a similar trajectory, but I do think there are some things that

make it trickier. Partly, it's trickier because it's a more complicated technology. Partly, it's trickier because it's moving faster. And so the sort of question of, how often do we have to update our ideas about what the appropriate safeguards are, and how often do we have to update our ideas about what the threat landscape is, are pretty significant. I still think that there's potential here for both regulation and insurance to play an important role, but that there needs to be some attempt, as there has been in cybersecurity over the past several years, to require more reporting so that we know what these threats look like, so that they're not just happening internal to companies or individual users, but they're being collected and analyzed somewhere. And I think the other lesson of cybersecurity policy that's really helpful is that we don't want to wait and leave it to private companies to figure this out for themselves. Which I would say was the approach in cybersecurity for many years, until the incidents reached the point of a fuel pipeline has been shut down because of a ransomware attack. And so as the sort of stakes of the incidents got higher, you saw more and more policies coming out trying to deal with cyber threats. And I think in AI, it would be wise back to the question of, is there a catastrophic risk on the horizon not to wait for a catastrophe to start thinking about some of the reporting and safety requirements.

Michael Klein
I like the idea of not waiting for a catastrophe, And, I also, I want to mention in closing that I did not use Chat GPT or any other large language models to come up with these questions, and I'll leave it to the listeners to decide whether I should have done that in fact. But I'm sure whatever the questions, your answers, Josephine, have helped people understand this timely and important issue. So thank you very much for taking the time to speak with me today on this topic.

Josephine Wolff
Thank you so much Michael.

Michael Klein
This has been EconoFact Chats. To learn more about EconoFact, visit www.econofact.org. EconoFact is a publication of The Fletcher School at Tufts University. Thanks for listening.