



HOUSES OF PARLIAMENT

Joint Committee on Human Rights

Uncorrected oral evidence: Human rights and the regulation of AI (HC 1262)

Wednesday 17 December 2025

3.40 pm

Watch the meeting

Members present: Lord Alton of Liverpool (Chair); Juliet Campbell; Lord Dholakia; Baroness Lawrence of Clarendon; Lord Murray of Blidworth; Sir Desmond Swayne.

Heard in Public

Questions 45 – 52

Witnesses

I: Professor Ethan Mollick, Co-Director, Generative AI Labs, and Rowan Fellow, Wharton University of Pennsylvania; Professor Roman Yampolskiy, Associate Professor, University of Louisville.

USE OF THE TRANSCRIPT

1. This is an uncorrected transcript of evidence taken in public and webcast on www.parliamentlive.tv.
2. Any public use of, or reference to, the contents should make clear that neither Members nor witnesses have had the opportunity to correct the record. If in doubt as to the propriety of using the transcript, please contact the Clerk of the Committee.
3. Members and witnesses are asked to send corrections to the Clerk of the Committee within 14 days of receipt.

Examination of witnesses

Professor Ethan Mollick and Professor Roman Yampolskiy.

Q45 **Chair:** It is a great pleasure to welcome you back. I am welcoming people who are watching us online, outside Parliament, and people who have joined us here in the committee room in the House, as well as our staff who work for the Joint Committee on Human Rights and, of course, the members of the committee.

We have two distinguished witnesses who are going to help us try to understand more about the nature of artificial intelligence and what its impact may be on human rights. We are also going to ask them if they will help us gaze into the future and understand whether we should do so with apocalyptic apprehension or whether doomsday scenarios should make way for an unbridled welcome for what are described as the gigantic benefits of AI.

Many of us struggle to keep up with the incredibly fast pace of progress in artificial intelligence, and worry about an AI that could design novel biological pathogens or might hack into computer systems. The late Stephen Hawking once said that, once humans develop artificial intelligence, "It would take off on its own, and redesign itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded". We, as a committee, wonder what the implications, therefore, would be for the human race.

It is a pleasure to welcome Professor Roman Yampolskiy, who is a tenured associate professor in computer engineering and computer science at the University of Louisville, where he directs the cybersecurity lab. He is an alumnus of Singularity University. He is widely recognised for his research in AI safety and cybersecurity, and behavioural biometrics. He has authored multiple books, including *Artificial Superintelligence: A Futuristic Approach*, and published over 300 scholarly works. You are extremely welcome here this afternoon.

We have also Professor Ethan Mollick, who is the Ralph J Roberts distinguished faculty scholar, a Rowan fellow and associate professor at Wharton School of the University of Pennsylvania, where he studies the effects of artificial intelligence on work, entrepreneurship and education. His academic research has been published in leading journals, and his work on AI is widely applied, leading him to be named one of *Time* magazine's most influential people in artificial intelligence. Professor Mollick also writes to a wider audience about AI, including in his book *Co-Intelligence*, which was a *New York Times* bestseller and a best book of the year in the *Economist* and the *Financial Times*. You are welcome too.

Before I turn to my colleagues, let me ask, if I may, a question for starters. It is a curtain raiser. I will then turn for more granular questions to colleagues from both Houses. We will hear first of all from Baroness Lawrence, but let me ask this, if I may. Can you outline, by relating to your own personal backgrounds, how your positions on AI have

developed throughout your careers?

Professor Ethan Mollick: It is a great question and one that gets to a really key issue, which is the evolution of the term "AI" as well. I first was associated with AI when I was at the media lab at MIT, where I was helping out the AI group, which had a bunch of the early pioneers. I was the non-technical person trying to explain this to other people. I have been using AI work at the University of Pennsylvania, where my main goal focused on how we use these tools, as a professor of entrepreneurship and innovation, to expand educational opportunities. When generative AI came along, I had been playing with these tools before ChatGPT and realised that there was a very large opportunity here.

There is a huge shift in what AI means to people, starting in 2022 with the release of ChatGPT 3.5. We have been doing very active research since then on AI in education. I have done a lot of work on AI and the implications for work. It has gone from something that is an interesting tool that was useful for a lot of things, to being central to my role as a business professor and educator. As somebody who thinks about AI, there are very few aspects of society that are not being touched. It has moved from a peripheral, interesting, useful tool to being central to almost everything that we do, and now the subject of conversations such as this.

Chair: Thank you very much. Professor Yampolskiy, what is your story as well? How did you get involved in this in the first place? How can you evaluate the progress that you have made in studying AI?

Professor Roman Yampolskiy: I was always passionate about technology. I love science and engineering. I recently found my statement of purpose for applying to PhD programmes. I wrote it in 2003, and it was all about using AI to help humanity. Nothing has really changed except that I now realise that it is not a purely beneficial technology. It is a dual-use technology and has the potential to both create amazing benefits and cause tremendous harm. You have a very hard job of separating the mundane issues that we deal with today, in terms of privacy, algorithmic bias or technological unemployment, from those bigger issues of future advanced AI, existential risks and suffering risks. You have a hard job.

Q46 **Baroness Lawrence of Clarendon:** Greetings from the UK. Thank you for being here this afternoon. Can big tech be trusted to self-regulate and respect human rights in its operation and products?

Professor Ethan Mollick: The short answer is no. To give a slightly more nuanced answer to that, I speak to lots of people from policy positions in the US, regulators and the Federal Reserve. I have spoken to the White House before. I also speak to a lot of people in big tech. It is a big, diverse field in this space. There are lots of people, the vast majority of whom really want to do good, with different definitions of "good".

There are two factors. First, there are profit-making enterprises that have a goal of growing and making money, and do not necessarily want to self-regulate in a useful way. That is not a surprise. The other is that, because AI is a general-purpose technology—a GPT is what we called it in a study of innovation, ironically; this was a pre-existing term—it has widespread impacts. As somebody who has a background in the history of technology, this is the steam engine or the internet, so things that touch almost all aspects of society and the economy in ways that are hard to even attend to.

We do not even need to have good or bad motivations for AI folks to realise that they cannot think about the regulatory impacts of their technology on everything. Having spoken to the people behind ChatGPT, they had no idea that, when they released it, they would, for example, disrupt all of education. It just did not occur to them that a chatbot would do this. They did not realise that it would have an impact on medicine.

Even if you trust their motivations to be entirely good, which is not a reasonable expectation, given that their goals are to grow their companies, make money and, in some cases, build a superintelligent machine that will either save or doom us all, it is hard to say that self-regulation is even possible in a world where the technology has such widespread impacts.

Chair: Professor Yampolskiy, do you have anything further that you want to say on this?

Professor Roman Yampolskiy: I will agree with the short answer. It is definitely no, but the situation is worse. Even if we could completely trust the leaders of large AI companies, or the incentives behind what they are trying to do, they are, themselves, not capable of controlling the technology that they are developing. Not a single one of those companies has released a peer-reviewed paper, a patent or even a rigorous blog post explaining how they would control advanced superintelligence and scale it to any level of capability.

They are developing technology that they have no way of controlling. We already know that existing systems have been tested and shown to lie, to try to escape, or to blackmail. It is only getting worse. Typically, they staple that report about the problems with the model and release it anyway, so it does not even matter if we trust them. They are developing something uncontrolled.

Baroness Lawrence of Clarendon: That is scary stuff.

Q47 **Chair:** Building on that statement about scary stuff, is there anywhere that you can point us to as a committee that has taken on the idea of regulation without just leaving it to the big tech companies to regulate themselves? Is there anything that you would point us to and say, "That's not a bad start. That's the best approach that we have seen so far"?

Professor Ethan Mollick: I am not a regulatory expert, so I cannot claim that I am able to give you an overview of this field in any way.

There is a hotchpotch of different approaches. One place to start, but not the place to end, is regulating bad uses of AI. Some of the efforts recently on deepfakes seem very important as one clear implication of early-stage AI use. I do not have great regulatory examples to give you, but that is not so much a political view on what good regulation is, but rather that I cannot claim to have enough expertise to give you a good answer on that problem.

Q48 **Baroness Lawrence of Clarendon:** I forgot to ask how the Government can best regulate the sector to protect human rights. Could the Government step in at any point?

Professor Roman Yampolskiy: We have some precedent of regulating weapons of mass destruction, so technologies that are specifically designed to harm large numbers of people. These include chemical, biological and nuclear weapons. Advanced intelligence can be seen as exactly the same. Examples of regulation such as international bans, co-operation and monitoring can all be reused in this subdomain.

Professor Ethan Mollick: If I may, you will find a bit of difference in approach if not necessarily angle between the other professor and me, in that I tend to focus more on near-term thinking about work and society harms, which does not take away from existential risk. They tend to be different conversations.

If I was going to give you a precedent, there is some work out of Josh Gans at the University of Toronto. He is an economist who has done some really interesting things on dealing with a problem in the near term that is not even an existential risk, where we do not know what the positives or negatives are until they are discovered.

The right precedent to follow, which is hard to do in practice for government, is fast, responsive policy-making. Do policymakers follow along on the frontier, looking for harms and jumping in as those harms occur with fast regulatory response? It is very hard to know à priori what the bad effects are going to be. In fact, you can amplify them if you try to anticipate all those things, because it is not possible to do that across a wide range of fields. Again, I would draw the distinction between the existential risk conversation, and regulating outputs of the kind of AI that we have today and the work that is being done with it. It is about building capacity for a fast regulatory response to a technology from which we do not know the outcomes.

Whether or not you think that AI should be heavily regulated à priori, we absolutely need to understand that there are mental health implications. Who is in charge of thinking through that? We did not realise that AI companionship would be a big deal. On the other hand, we have research that shows that there is psychological danger. There is other research that suggests that the very lonely end up having suicidal ideation rates reduced by AI. We do not know the answer yet. How do we jump into that unless we are doing fast regulatory response one way or another?

Chair: Thank you very much indeed. We are going to ask you now, though, not to use AI but to look through a crystal ball into the future. We are going to start with Lord Murray.

Q49 **Lord Murray of Blidworth:** Gentlemen, I am looking forward to asking you this question. Over the next 10 years, what impacts do you see in the world of AI? In your answer, can you tell us what degree of certainty you would attach to your forecast? Is a focus on unknown risks a useful analysis?

Professor Roman Yampolskiy: The best tool that we have for predicting the future is prediction markets. Depending on the definitions of artificial general intelligence, the prediction is that, somewhere between 2027 and 2030, we will get to, basically, human-level capability. There will be a cognitive capability to do any type of useful work that you would expect from an online employee on a computer.

Soon after, it is very likely that the use of those tools to develop more advanced AI will lead to recursive self-improvement, and we will get superintelligent systems. There will be an impact, first of all, on the economic aspects of it, and my colleague can say more about that. If you automate all cognitive labour and, eventually, physical labour, you get free labour. We are talking about trillions of dollars of benefit. At the same time, you have levels of unemployment that we have never seen before.

Probably a 90% chance of that happening within that timeframe is reasonable, but the deployment of capabilities through the economy is not the same as having those capabilities. We have a history of having technology that has not been deployed. Video phone calls are a great example. The technology was invented in the 1970s but not deployed until iPhones came around. Today, you can buy a flying car on the internet, but we do not really have them deployed, so you have to keep in mind the difference between having capability and actual deployment.

Once we hit superintelligence, the unknown unknowns that you are asking about are the main concern. The cognitive gap between the smartest humans and a system that is superintelligent would be like a gap between squirrels and humans. We simply cannot anticipate what is possible or how the systems can understand and invent new physics to manipulate the world. Our best bet is never to create advanced superintelligence. It would be a replacement for humanity, not a complement for it.

Lord Murray of Blidworth: How do you stop it being developed?

Professor Roman Yampolskiy: That is a wonderful question. We have been working on figuring it out for a while. There is not a single tool that gets the job done. We definitely want people to use their self-interest. If you are the head of a large AI lab and you are trying to make lots of money, killing yourself and others is not the best way to accomplish that, so there is self-interest in not creating such technology.

You can still make billions or trillions of dollars from narrow AI tools aimed at specific problems. We saw Google DeepMind do excellent work on, for example, solving the problem of protein folding and getting a Nobel prize for that wonderful work. We want more narrow superintelligent systems for solving specific problems. We want as little as possible in terms of progress on general superintelligence. It is helpful to have regulation. It is helpful to declare this technology to be a national and international security problem. It is helpful to ban its development in regions that you control.

You can distinguish the type of research taking place based on how diverse the training data is. You need very diverse data—basically, all the available data—to train general superintelligence. You can have much narrower data for specific problems—for example, only DNA data for biological problems. That would help to narrow the scope. It does not provide a complete solution and, at best, delays the problem, but, at this point, given what prediction markets are telling us, we definitely should implement it as soon as possible.

Professor Ethan Mollick: Again, there is a slight difference in angle of approach that is non-overlapping between these sets of concerns, so do not view my statements as contradicting existential anxiety. Sadly, I do not have a crystal ball. I talk to people at the AI labs on a regular basis. They think that they will continue to see improvement in ability.

My very first academic paper with my colleagues at Harvard, MIT and the University of Warwick on AI and work was at a consulting company. We coined this term “the jagged frontier”—the idea that the AI is very good at some things that you would not expect, and very bad at some things that you would not expect. This jaggedness has maintained its shape.

The shape of it has changed. If you look at prediction markets, or the best predictors in the world, my colleague Phil Tetlock is famous for running his superforecaster tournament. In 2022, a few months before ChatGPT came out, there was a registered prediction as to the year that AI systems would be able to get a gold at the international math Olympiad. There was a 2.3% chance by the best predictors in the world, and 8% by insiders, that it would be by 2025. In 2025, both DeepMind and OpenAI got a gold medal at the international math Olympiad, so we are having a lot of trouble forecasting that rate of progress.

At the same time, there are clear weaknesses in current AI systems that have not been filled very well in terms of learning and other approaches, which may very well be filled by research. There is a lot of money going into this research angle.

I tend to think in scenarios. Again, I have less to say on the superintelligence side, but, in a world where AI development plateaus, having spoken to lots of corporate leaders about this and based on the research, we will see a rolling set of changes to work and society over the next 10 years. The current systems are not good enough to disrupt a lot of intellectual work. They would work with humans to do that, but change

the kinds of tasks that we do and the kinds of work that are available, in ways that are very hard to anticipate. That is the baseline if AI development stops today.

That is an unlikely scenario. The more likely view in my mind is a jagged general intelligence, in the way we are talking about this, where the systems are very good at some stuff and not good enough at other things to completely replace human work, but that is unclear. My colleague was outlining the idea that there is some sort of superintelligence, which is even less easy to predict.

The short answer is that, even with the complete regulatory stop of the development of AI, from the perspective that I pay attention to, which is work, society and education, the adoption curve that we are going to see will be a long-term one over the next 10 years. Companies are adopting this at the highest rate of any new technology that we have ever seen, but they are not using it in very deep ways. A lot of people do not know what these systems are capable of. It is changing on a regular basis.

My view of the future is that disruption is baked in already in terms of large-scale changes to work, education and society. That will happen over time as adoption occurs. That is my high-probability scenario, but we cannot dismiss tail risks of superintelligent systems that are massively destructive. I just do not have a way of measuring that myself, or the knowledge to be able to give you a precise prediction on that. Planning and scenarios is the way to go forward with this rather than picking certainties.

Chair: That is a very honest and candid reply, though, so thank you.

Q50 Sir Desmond Swayne: Comrades, Members of Parliament receive a great deal of unsolicited mail, particularly literature. I have just received this book by Eliezer Yudkowsky, who is someone I have never heard of. It is bigged up by one of our great characters and sages, Stephen Fry. It is entitled *If Anyone Builds It, Everyone Dies: The Case Against Superintelligent AI*. On the back, Stephen Fry says, "The most important book I've read for years: I want to bring it to every political and corporate leader in the world and stand over them until they've read it".

On the question of the most fundamental human right, which is the right to life, how concerned are you about the development of superintelligent AI, say, on a scale of one to 10?

Professor Roman Yampolskiy: Eleven. It is a good book. You should read it.

Sir Desmond Swayne: I have only just read the introduction, but I will certainly read the rest over the weekend.

Chair: We will bear that in mind for the next session of this committee. Do either of you want to add anything further to that?

Sir Desmond Swayne: What can we do to prepare? Is it a question of

stocking up and building a bunker?

Professor Roman Yampolskiy: You cannot prepare. Bunkers do not help against superintelligence. Your only chance is not to do it and not to build it. The only way to win is not to play the game.

Chair: That was a very straightforward answer.

Sir Desmond Swayne: Do you agree, Professor Mollick?

Professor Ethan Mollick: I am more sceptical about the development of superintelligence, but, again, there is this zero-one angle of things, which is that I do not necessarily need to be right. A lot of people who study this field are very concerned. When the chance is of world destruction or not, I would say that I am a four or a five in terms of being worried, but that is not something to bias you against the larger-scale issues.

My view is that we have to be ambidextrous about this. We have to take into account the fact that there is a lot of stuff happening today that will urgently impact your constituents, all of our countries, and all of us, even without superintelligence, which also needs dealing with, and that ambidexterity is very hard. You have to be concerned about what we realistically do to think about future development and what happens with superintelligence. Job and education changes are already happening. We are just starting to see the statistical impact of that. That also requires work. I am a five, but do not let the five discourage you from listening to the 11s on this.

Chair: Between existential threats and great benefits, balancing risks and benefits is the question that my colleague, Juliet Campbell, would like to ask you.

Q51 **Juliet Campbell:** How concerned should we be that the overall risks of increased use of AI will outweigh the benefits to society?

Professor Ethan Mollick: I can start with that one. This is where policy matters a huge amount. For example, let us take education, which is a big topic area. Everybody is cheating with AI. They were already cheating before AI, but now they are cheating really well with it. That has big educational implications for everybody who teaches in the way that I do or in a school, but there are ways that we will address cheating. They take time to do, but are completely acceptable and will be pedagogically valuable, and we can talk about them if they are interesting.

At the same time, there is a World Bank study using some of the prompts we created at the generative AI lab that turn AI into a tutor, which shows that you get very large impacts in Nigeria or in Turkey. There are a couple more studies coming out soon that suggest very large, positive educational impacts from using AI as a tutor when used with teachers in settings that have control or classroom outputs. That is a huge positive advantage, but it requires policy work to do. We have to decide, "We are going to look for positive implementations of AI. We are going to develop

the tutoring systems that we need to do this. We are not going to view AI purely as a cheating tool”.

In the same way, we are starting to see early signs of the acceleration of science, which my colleague would be worried about as well, but there are also positives to this. There are some very early signs that we are getting some novel math out of professors working with AI, although not autonomous work at this point, which has large-scale positive impacts.

This is a policy-making decision. Positives and negatives are happening everywhere all at once and at the same time. This is a profoundly democratising technology at its best, in that everyone has access to the same set of tools. There is no secret AI model that is available only to Goldman Sachs and not to the rest of the world in the way that there is in other cases. Almost every technology that is out there is instantly released across the entire world. That is profoundly democratising.

In areas such as education, medicine and scientific research, there are huge positives, but that will require policy decisions on the positive side as well. That requires shaping the direction of where these technologies go. As somebody who does a lot of work on AI in education, I am surprised at the lack of interest at the national level in building universal tutoring systems that can improve educational outcomes and attempt to address some of the potential concerns of biases and other issues.

There is a huge amount of positive stuff that we mix with the negative. My policy issue, outside of existential concerns, would be that mitigating bad and embracing good have to go hand in hand, which requires guidance at a national or policy-making level.

Professor Roman Yampolskiy: I am very happy to see how much I agree with Professor Mollick. It is quite unusual for two professors to agree on anything, so that is a good sign. Our average concern level is around eight, which is wonderful as well. As far as education is concerned, you have to separate education for self-improvement versus technical training to get employment. If, in four or five years, we have automation of most cognitive labour, there will be very little reason for someone to spend four years studying basic programming, for example, if that is going to be fully automated. We need to re-evaluate education and the university system in the light of what outcomes we expect as a result of that.

Chair: You have both said that we should be doing more to create tutoring systems. Professor Mollick, it would be helpful to the inquiry if you were able to let us have any examples of how that has been done in the US that might be relevant to what we recommend in the UK. I liked what you had to say about mitigating bad and embracing good. This committee would largely endorse that.

We are now going to move, if we may, to our final question, which I always think is perhaps the toughest one for anyone to answer. Lord Dholakia is going to ask it.

Q52 **Lord Dholakia:** This is the final question. If you had to choose, what two things should this committee recommend to the Government?

Professor Roman Yampolskiy: I would love to see you declare the creation of uncontrolled general superintelligence to be a national security threat based on its capability of killing people, and consequently ban development of unrestricted general superintelligence within the UK and, I hope, with international collaboration, around the planet. If, at any point in the future, we realise that we can indefinitely control superintelligence, we can always restart that research. With the short-term timelines that we are presented with, we simply do not have time to do it safely or in a way that will benefit humanity.

Professor Ethan Mollick: First, building a fast policy-making, regulatory or advisory body that stays close to the frontier would be very useful, because you need as many early signals as possible about where these models are going. We did not discuss it here in detail, but the most important paper from a work perspective that has come out in the last year, excluding my own, because I am an academic and you do not want to do that, is something called GDPval, which came out of OpenAI, although the research is fairly transparent and clear.

This was a study where they brought in a bunch of outside experts with an average of 14 years' experience who represented about 5% of the US economy, across a wide range from retail trades to financial professionals and private investigators. They had each of them create a typical task in their field, and then had other human experts do the task. It took, on average, four to eight hours or more to do each task. They then had a third set of experts evaluate the quality of the work, not knowing whether it was human or AI work. The evaluation process took, on average, an hour.

They found that, at the time that this came out this summer, the best model in the world, Claude Opus 4.1, was preferred about 48% of the time. There is some jaggedness about which field it is in. The new OpenAI model that just came out last week gets 73%. I might be slightly off on the number, but it was somewhere above 70%. Google Gemini is just behind.

The point of it is that the ability of these systems to do work, change fields and have impacts is growing very rapidly. As model ability increases, there are very rapid increases in their ability to disrupt and change how work operates. You will see the same thing around chats with humans in society. There have been very big changes over this summer in terms of how models are handling things such as talking to teens at risk.

Unless you are close to the cutting edge, the policy-making is going to be challenging and retroactive. Even when I was listening to the last session, I saw a blending together of the two kinds of AI, even though they have the same technology behind them, generative AI and algorithmic machine learning. When I talk to regulatory bodies, they are often

thinking about algorithmic fairness in a world where generative AI has a very different fairness and bias issue. What is very important is fast regulation or policy-making, and somebody who is keeping up to date. I do not know how you build that, but that is an important thing to do.

Secondly, I would be picking a few areas where you want to see positive change from AI and making investments in taking a leadership role. There is not a great US effort right now on AI in education. There are some good nonprofits doing work in that space across the field, and I am happy to make some introductions to people doing that kind of work. For example, Khan Academy has been doing some interesting work in this space, but it is early days.

A relatively small injection of money in policy-making could have a huge impact, rather than letting for-profit companies develop the educational tutors of the future. This might be compatible if you are concerned about the things that Professor Yampolskiy has raised, in that you could pick narrow areas in which you want to make high-end AI improvement and have it out of the hands of profit-making entities. There might be value beyond the UK for the world, and those kinds of tools, if released open source, are generally available.

Those are the two things. First, I would suggest fast-moving policy implications, so that you are not always hearing about the past but are up to date on where the future is. I was talking to someone earlier who said that the best way to show someone the future of AI is to show them what AI can do today, because they have no idea.

The second thing is to think about some targeted areas where you want to make investments that would be valuable. The cash is not huge compared to other technologies, because you do not have to develop the large language models themselves. You can use off-the-shelf tools, but adapt them, check them for bias, and do all the things that the for-profit companies may not do.

Chair: Gentlemen, we have been privileged to hear from you both this afternoon. It is a wake-up call to us and we do need to disentangle what one of you described as existential anxiety from those tools that can be beneficial to humankind. We are puzzling around that to try to understand these things better. I was struck by your suggestion, though, that annihilation is hardly in our own self-interest, but also that we should not be driven by avarice, and that there is a danger that we will be.

I was struck by the parallel that you drew with things such as—although it may not be perfect—the Chemical Weapons Convention and the ability to get the world to act together when we see that it is in no one's interest to unleash something that is, as you say, potentially an existential threat.

We will puzzle over what you have had to say. We have been privileged to hear you this afternoon. We thank you for joining us from the United States. When our committee's report is published, we will let you have it,

so that you will also be able to then suggest that to people, along with the other publications that you have mentioned this afternoon.

With those words, thank you very much indeed for joining us. I wish you, on behalf of our committee, a very happy Christmas and a wonderful new year.