

Treasury Committee

Oral evidence: AI in financial services, HC 684

Tuesday 24 June 2025

Ordered by the House of Commons to be published on 24 June 2025.

[Watch the meeting](#)

Members present: Dame Meg Hillier (Chair); Dame Harriett Baldwin; Rachel Blake; Chris Coghlan; Bobby Dean; John Glen; John Grady; Dr Jeevun Sandher; Yuan Yang.

Questions 93-130

Witnesses

I: Professor Sandra Wachter, Professor of Technology and Regulation, University of Oxford; Professor Galina Andreeva, Personal Chair, Societal Aspects of Credit, and Director, Credit Research Centre, University of Edinburgh Business School; Professor Neil Lawrence, DeepMind Professor of Machine Learning, University of Cambridge.

Examination of witnesses

Witnesses: Professor Sandra Wachter, Professor Galina Andreeva and Professor Neil Lawrence.

Q93 Chair: Welcome to the Treasury Committee on Tuesday 24 June 2025. We are here to continue our inquiry into AI in financial services. We previously heard from financial institutions that are using AI in their work, or considering how to bring it in. Today we have a panel of expert academics who have been looking into how this will work and raising some concerns about how AI may be used, so we are keen to hear their perspectives.

I am very pleased to welcome with us in the room today Professor Galina Andreeva, who is the personal chair of Societal Aspects of Credit and director of the Credit Research Centre at the University of Edinburgh Business School. Also in the room is Professor Neil Lawrence, who is the DeepMind professor of machine learning at the University of Cambridge. They are joined online by Professor Sandra Wachter, who is the professor of technology and regulation at the University of Oxford.

I mentioned that we had various people in from the finance sector, and UK Finance told us that the financial services sector is leading the way in the adoption of AI. Would you agree with that?

Professor Wachter: Yes. I feel that almost every sector is completely penetrated and transformed by AI; we can see that trend very, very clearly, especially in the finance sector. The Bank of England published a report last year when it found that 74% of all financial firms in the UK are already using some form of AI, and early this year, Goldman Sachs, for example, announced that it is going to roll out even more AI services within the organisation for bankers, traders and asset managers. But it is not just Goldman Sachs: J.P. Morgan, Morgan Stanley, Deutsche Bank and Moody's are all heavily investing into those solutions, predominantly still from third-party vendors like OpenAI, Google and Meta. You can see that almost every aspect of the financial sector is trying to embed AI as much as possible.

Professor Lawrence: It is natural because these technologies depend a great deal on data and in many areas that data is not available, but in financial services I often think of it as data with dollar signs or pound signs. A lot of the things that prevent the adoption in other fields—such as data quality—are easily taken care of in financial services because we already have very large mechanisms, in particular accounting, for managing data quality, because it is very clear what goes wrong when you do not manage data quality. That is a big difference to other sectors where it is not so directly clear what happens when you do not manage data quality. That is the only thing I would add to Professor Wachter's wonderful introduction.



Professor Andreeva: I agree. In the previous session with financial institutions, they made a distinction between predictive AI and generative AI. Predictive AI has been used by financial institutions for decades now, so it is not really a new process for banks; they have been using AI extensively in various functions and applications. One of the early success stories is the application of neural networks in fraud detection, and that goes back to the last century.

Professor Lawrence: To the 1990s, yes.

Professor Andreeva: Generative AI may look like a big novelty and breakthrough but from inside financial institutions, this is just the next step in the process.

Q94 **Chair:** Professor Andreeva, do you think that there are more risks with generative AI? We want to talk more about regulation later but overall, is there a different challenge to the technologies that banks or financial institutions have previously used?

Professor Andreeva: Good risk management always implies that we should be constantly checking the horizon for new risk. At the moment I see old risks amplified or brought to the fore with this new technology, but nothing fundamentally different.

Professor Wachter: I would agree that there are certain risks that we see with predictive AI that we also have with generative AI, such as bias, explainability, questions of data protection, questions of data access. Those are risks that we see with predictive AI—decision-making systems that help you decide and allocate resources—but they might manifest in a different way. Bias in generative AI looks different than bias in allocating resources. It might be a similar harm, but it might exemplify in a very different way, which means you have to have different detection mechanisms in place.

The second point is that some risks might be actually quite new. I am thinking in particular of the issue of hallucinations: that means that a generative model spits something out that is factually inaccurate. This is definitely something that we have to take seriously and is one of the biggest problems and risks when it comes to the adoption of generative AI.

Professor Lawrence: Is it okay if I set a little context about what I think the shift is? I want to be very clear: a lot of the talk we are hearing about artificial general intelligence is total nonsense, but there is an extremely significant shift to this technology and I can only put that in place by setting the context of digitisation in the first place.

I am here talking to you, and I am communicating with you in information theory terms at about 2,000 bits per minute—so about 2,000 coin tosses a minute. If I were a machine, I could share information with another machine at 600 billion bits per minute. To give you the sense of



the difference of that, it is the difference between walking pace and the speed of light. Two machines are sharing information with each other at light speed and humans are sharing information at walking pace.

That has actually led us to a bit of a dystopian environment already. Look at things like the Horizon scandal, where you have increasing disempowerment of individuals—in that case, politicians, lawyers, accountants, sub-postmasters—all in the face of a digital system that is operating much faster and is therefore much harder to understand, and responsibility is not being properly allocated to those who manage that system. What does that have to do with generative AI? Well, that underlying challenge, where there is this data ecosystem operating that much faster, is leading to a number of societal problems—whether that is social media or whatever—where there are new routes for manipulation.

What generative AI does is something quite extraordinary actually, because for the first time it means that literally anyone in this room can talk to a computer and have the computer understand what they are saying. I think the character was called Jo Hamilton in the dramatisation of the Horizon scandal, where she ends up at one point screaming at her machine, “That is wrong; you have just reconciled that incorrectly!” Of course, we all know the machine cannot hear that, but this technology means that machines can hear that. It offers an opportunity to be in a situation where, instead of having to spend years searching through software, we can actually see what the flaws and errors are quickly, re-empower the lawyers, the accountants, the politicians in that process and not have to wait 23 or 24 years to deal with what was fundamentally a software bug.

Parking the ridiculousness of the things we are hearing about artificial general intelligence, which broadly speaking are quite eugenic in origin, we can see that that is an enormous transformation. It is shifting away from the current dystopia, where those who can control digital systems are highly empowered and those we expect to be responsible in society are highly disempowered, which is most starkly shown in the Horizon scandal, but we can see that across society.

Chair: Brilliant. Thank you very much indeed.

Q95 Chris Coghlan: Professor Lawrence, Citigroup has warned that the banking industry would be hardest hit by the deployment of AI, estimating that 54% of roles are at risk. Do you think AI is going to increase or decrease employment?

Professor Lawrence: It is a fascinating question, because when they mean hardest hit, they are referring to their financial model. The interesting question we are very often missing in these conversations is: what will it mean for customers and citizens? We face challenges around existing regulatory environments: historically their way of interfacing with the wider environment is typically through businesses directly rather than through individuals who are affected. But as technology moves very



quickly, what we need to get much more focused on is how this is affecting the customers of those banks and the citizens in society, rather than the financial models of those banks themselves.

Yes, I am sure they will be disrupted, because of the reasons I just outlined. The question that I am interested in is: how do we steer that disruption in a way that is not economically so disruptive that it is changing things too rapidly, which is problematic, so that the destination we reach is one where the citizens and customers in society are better served?

Professor Andreeva: There is a risk of very rapid change and disruption, but I believe there are processes of adaptation in place. In general, financial services are very conservative, so I would not expect the banks to jump and embrace the new technology without thorough checks regarding all the implications. However, regulators should also be part of the process, and maybe this particular process should be closely monitored just because of the speed of change that we are witnessing right now.

Q96 **Chris Coghlan:** Do you see that the productivity gains in AI could lead to greater wage disparities between high and low-skilled workers?

Professor Andreeva: I am not sure if I would expect this to happen. I would rather agree with the opposite sentiment: that AI has a huge potential—generative AI in particular—by levelling the field and allowing non-technical people to embrace this new technology and actually learn: learn coding, learn programming, learn how to use technology for their benefit.

Q97 **Chris Coghlan:** Professor Wachter, what do you see the impact of AI is likely to be on people who are under-represented in the financial services workforce, such as women or people with disabilities?

Professor Wachter: The issue of workplace automation is one we have to take very seriously. If you look at the report from Goldman Sachs two years ago, in 2023, it said that it thinks that roughly a third of its own financial services workforce will be automated. The report even says that roughly 300 million people could lose their jobs worldwide. There are also staggering reports from McKinsey that suggest that between 400 million and 800 million people could lose their jobs due to the implementation of AI. What statistics are actually correct is hard to say—it is really hard to predict the future—but what we know already is that one of the selling points of AI is to reduce labour costs. If you go to the financial service industry conferences, this is definitely one of the selling points. There will definitely be an attempt to automate, at least partly, some workforce away.

As Professor Lawrence said, this is something we have to be aware of. This is a trend that is coming for better or for worse, and we need to figure out how we deal with that: what we do with the folks who are at



risk of losing their jobs and what it means for a society, in general, if you are losing the interaction with customer services and you are no longer talking to people, basically constantly talking to an AI system, and what that means. We have to take this very seriously.

In general, the AI automation of the last couple of years—especially in the tech sector—has seen so many layoffs: there were 50,000 layoffs this year in the tech sector alone. It is a trend, it is definitely declared a goal and we really need to take this seriously. As always when it comes to firing waves, those that are already disadvantaged in our society are usually the ones who lose their jobs first, including women and people with disability. They will be at greater risk of losing their jobs first.

Q98 Chris Coghlan: With any new technology you both destroy jobs and create them. For example, the car created the whole motorway services industry and tarmac roads. Is it too early to say whether AI is going to be net job creating or destroying, or we just do not know at this point?

Professor Wachter: I think it is not like the Hydra where you cut off a head and two just spring in its place. I do not think that is how job creation works. Usually that means active investment, and when we want to create new jobs, this needs to be like a political will that comes from that. It is about really thinking about what we can do to incentivise companies to create new jobs because, in general, if you are in the business of gaining productivity, you are not in the business of creating new jobs unless you have to. We have to find incentives for companies actually wanting to create new jobs.

If you think about the industrial revolution, at the beginning we had 40% of the workforce working in agriculture; now we have 2%. The reason why we did not face mass unemployment was because Government stepped in and made sure that people were protected. A similar strategy needs to happen right now.

In the immediate term, it is not as important to think about how many people are going to lose their jobs. We already know that people will want to pay their rent, so we need to think about what kind of jobs spring up in its place. Is this just going to be gig workers? Are they going to have the same protection as normal employees? Are the wages going to be comparable? Are they going to have firing protection? It is really about making sure that the new jobs that are created are not worse than those we already have.

Professor Lawrence: It is very difficult to say what is going to happen, and one of the problems we have at the moment, particularly in the UK, is highly confident people claiming what is going to happen. The one thing we know is that highly confident people's claims of what is going to happen are not going to happen—and I am confident about that.

But we can say a few interesting things. One is that we do not expect humans to stop being interested in other humans, right? Whatever the



future economy is, there is going to be a lot of interest in humans being interested in other humans.

Another thing we can say is that the quantity of human attention is not going to increase. The bottleneck in the future economy—indeed even in today's economy—is what you might think of as the attention economy, where it is all about our limited attention as human beings. That is the second thing you can say.

The third thing you can say is that businesses will not stop trying to corner that market. You will see businesses trying to capture human attention. That is what we have at the moment—what I think of as the attention capture cycle—with businesses like TikTok, and that is the best way to monetise.

The challenge from the perspective of the public sector, the Government and so on, is how we steer that in the right way to get the sort of outcomes we want for our citizens, not just for large businesses. Unfortunately, the balance has gone in the wrong direction for the last few years. But we have a spectrum of possibilities in the future where, if we could get it to a place where the empowerment is going back more towards the individual, smaller business and medium enterprises—the sort of backbone of the UK economy—that would be a much better place to be.

Q99 Yuan Yang: Professor Lawrence, could you expand on your point that over the last few years the trend has gone in the wrong direction?

Professor Lawrence: One of the starkest examples of it is the Draghi report in Europe, which is an interesting report because it is mainly written by macroeconomists. I guess the headlines are that the US has grown faster than Europe—we can see the UK as just another European country within this context—and that growth has all been in the large tech companies.

There are two ways of viewing that. One is that we therefore have to have large tech companies, because otherwise we are not going to grow as fast as Europe. Or the other is that there is something very strange going on, because when we say large tech companies, we are not talking pharmaceuticals—we are not talking about the type of technical industries that are affecting people's health and improving our lives; we are talking about large conglomerates that, arguably, have a monopoly around our data.

There is work from colleagues at UCL, and they would call it algorithmic rent seeking. Their argument for what is going on here is that those small sections of the economy—very large companies, but a small section of the economy—are taking all the growth out of the areas that we would like that growth to be in, which is smaller businesses that are perhaps servicing local teachers or local nurses, and actually addressing the areas that we know our citizens care about.



That is the sort of headline trend, and then there is other really interesting work on this: my colleague Professor Diane Coyle's recent book *The Measure of Progress*, about whether the measures we are using for our economy around things like GDP are actually capturing this trend correctly. Consider the challenge we have had at the Office for National Statistics recently where it is struggling to capture basic measures like inflation. This is an immense challenge. This very weird thing is going on that these classical statistics we used to rely on are becoming less reliable, as is our ability to access them. As Diane points out, we cannot get the statistics out of these large tech companies in terms of how money is moving. During the pandemic, we were having to pay for data about our own citizens and how they were moving. So there needs to be a major change in the way we think about how we are monitoring our economy, because I do not think it is the ONS' fault per se: this is a dynamic shift in the way data is generated.

Q100 John Grady: Professor Lawrence, to pick up on some of your earlier comments, I would like to explore the core task of financial services regulation in the context of AI. One of the core tasks of regulation here is to regulate against overconfidence, animal spirits and the punch bowl. You said that artificial general intelligence is an absurd idea, and Professor Emily Bender has described AI as "a stochastic parrot" pointing to the fact that she believes the technology is overhyped. Does it follow, from what you say and what Professor Bender has said, that one of the real risks in financial services is that we are placing excessive confidence in AI, and then we will find it will not solve the problems it was designed to solve and, in fact, its overconfident deployment creates serious problems further on?

Professor Lawrence: Yes. It is fascinating because I am immediately thinking of what financial services are trying to do, which in the end is supposedly hedge, but individual people will tend to be overconfident. We see the same thing in our scientists: science as a whole tends to get sensible answers, but individual scientists tend to be overconfident. Yes, there is a real risk at the moment, and to highlight Professor Bender in particular and her colleagues, her colleagues were fired from Google for warning of these technologies a year before they came on the radar of this Government and the Government went into a mad panic.

To come back to what we were talking about earlier in terms of inequality in the workplace, what we have seen is—certainly in the UK advice infrastructure—female voices being almost eliminated. They have gone from being very omnipresent four or five years ago in terms of advising the Government on this, to basically being removed and replaced with male voices from either the tech sector or the entrepreneurial sector. This is a real warning, and my favourite reference on this is Claire Craig's book *How Does Government Listen to Scientists?*. We need honest brokers giving the advice here; we do not want issue advocates.



The exact phenomenon you are talking about is one when issue advocates are dominating the agenda. Over the last two Governments we have seen a lot of sensitivity to that emerging. Because of their confident voice, what I call type 2 stochastic parrots—Emily is talking about the machine; I am talking about human beings who know very little but speak confidently, like a fleshy chat GPT—are unfortunately tending to dominate advice at the moment. That will undermine the need to hedge that you are talking about, not just in financial services but across a range of areas.

Q101 John Grady: You make an important point. Professor Wachter, this traverses grounds you have discussed in terms of the nature of consultation and how regulation is formulated. Does it remain your position that legislators and regulators are not talking to all the right people here?

Professor Wachter: The most important thing about this is something similar what Professor Lawrence said: you need to understand how technology works to understand how it is disrupting the law. With our large language models, even though they are presented as those magical beings almost, or magic eight balls that tell us something about the truth of the world, what they actually do is just try to predict the next word in a string of words. This is why Bender and her colleagues call it stochastic parrots: it is just parroting back what it has heard.

But we take it very seriously when we hear it, and that is not coincidental because those systems are designed in a way that they are very convincing, which again is problematic. We tend to think of them as these knowing beings when in reality they are more or less like if you are texting with a friend on WhatsApp and you just press the middle key for a little bit, it will pop up next-word suggestions for you. That is what a large language model does, but it does it in such a confident way that all of a sudden we believe that it actually says something that is knowledgeable.

This is where the risk comes in, right? If you do not care about detail, if you do not care about truth, you can use a large language model. But in areas where truth and detail matter—the financial sector being a prime example—it makes a big difference if something is a two or a four, if one word is missing, or if you are comparing reports and making suggestions on how to invest in the market. It makes a big difference if there are mistakes in there, but it cannot tell you, “I don’t know,” and that is the biggest crux of it. If you want to think of it as a mental model, I would never think of it as a magic eight ball that tells you truth; it is more like a somewhat unreliable research assistant who is trying to please you and cannot say, “I don’t know.” So whatever you see, you always have to take with a grain of salt.

When implementing these systems, they can be very beneficial for the financial sector if you are doing things that are low risk, and if you are automating certain things where the processes are good. But as soon as



you are asking something about the future or truth, you are getting in muddy areas, because the training data is biased and the technology has no understanding of truth: it is just trying to predict the next word in a string of words based on what it has seen before. That means we have a system that suddenly spreads out lies, and the Government now need to think about how we deal with a liar that has no motive, but just lies for reasons that do not really exist. We need to think about the current regulatory structure and whether we have enough safeguards in place to protect asset managers, to protect customers and the economy as a whole, and so that these systems are not destabilising the market.

Q102 **John Grady:** Because the clock is running on, could you each name three groups that you think are under-represented in the discussion about AI, and three topics I should be thinking about in terms of what needs to be regulated in AI and financial services? I'm afraid I am not offering a prize for the shortest answer, but I would appreciate shortness.

Chair: Do not worry if you cannot hit all three on both—it is not a test.

John Grady: It's not a test; it is just literally three words for each: under-represented groups and things that we should be talking about regulating.

Professor Andreeva: Thank you very much for this question. Under-represented groups in discussion of AI are technical specialists, and especially specialists in maybe not mainstream applications of AI. When we talk about generative AI, the majority would think of investment banking as the main application in financial services, but financial services are not restricted to investment banking. My expertise is in credit risk; I would totally support more inclusion of credit risk specialists and risk professionals generally from all the different sectors and directions that financial services cover. That is my answer for under-represented groups.

On under-represented topics, I raised bias in my written evidence. When it is discussed, it is misunderstood, and maybe there is a need to bring in more diverse parties, such as academics, and not just technical specialists but social scientists, consumer protection groups, bankers, decision-makers and regulators to discuss this issue.

John Grady: Thank you very much. Professor Wachter, what are your three under-represented groups and three issues we should be looking at? Please be as short as possible—imagine I have the attention span of a gnat, which is correct.

Professor Wachter: Okay, I will try my best. In terms of under-represented groups it is definitely really important to listen more to independent scientists and less to people who have a financial interest in this technology being deployed. That is a big issue. We also know that voices from women and people of colour have been eradicated over the last years, even though 10 years ago they were leading the charge with



HOUSE OF COMMONS

showing the risks, benefits and ethical conundrums of AI. Those are unfortunately silenced at the moment. So more people from universities, people of colour and women—those are the groups.

In terms of areas, one that nobody really wants to talk about is the environmental impact of AI and the resources that it costs. This is the actual existential risk, if we want to think about that in those terms. The danger that causes to humanity cannot be overestimated. The environmental impact bias—as already mentioned—was a problem back then, is a problem now and will be a problem in the future. It is one we have to take very seriously.

The last one is hallucinations. I do not necessarily mean the type of hallucinations that you can immediately spot when it says something wrong, such as, “Put glue on your pizza to prevent the cheese from slipping off”—those things we can spot. I mean the more subtle types of hallucinations where people are not immediately able to detect it. For example, Stanford researchers found out not too long ago that legal tools that are being used to give legal advice are wrong 60% to 80% of the time. Nobody would hire that type of lawyer, yet we are implementing those systems, and the same will come true for the financial sector. Hallucinations are really a core problem that we have to take very seriously.

Professor Lawrence: Just a clarification on my previous answer: regulators are doing an extraordinary job at convening, but they are extremely under-resourced. In particular, it was appalling when we gave £100 million to a safety institute on the back of international warnings of killer AIs and only £9 million to our regulators to adapt. We still have not rectified that. Where there are problems in regulators, a lot can be due to under-investment and using them as a whipping boy for lack of innovation, which is a deeply problematic narrative.

On your under-represented groups, I would say teachers, nurses and habitual gamblers—the sort of people who are gambling on their phones the whole time, and their integration with financial services is through various well-known gambling companies that are constantly advertising alongside the football. These are the real people in our country that are affected by the decisions we make at these high levels. Those effects can be very significant, because those people are often on the margins and they do not tend to be represented in the economic statistics very well, because we cannot get a strong macroeconomic signal from them as people who have particular demands—demands to reduce the amount of time they are spending on data entry or whatever. Now, I appreciate that may be a little outside the direct remit of the financial services, but at the end of the day that is what we are trying to enable with finance, right? It is not just about the numbers; it is about better lives for our citizens.

Q103 **John Glen:** Professor Lawrence, could I clarify what you meant in terms of the regulators’ budgets? My understanding is that the PRA and the



HOUSE OF COMMONS

FCA, in financial services, for example, are funded from levies. They determine their own budgets on a business plan and so on and are scrutinised by this Committee and others.

Professor Lawrence: That was additional funding; it was when the £100 million was given. How long ago was that? I think it was two and a half years ago. A new security institute that I do not think has a regulatory remit was given £100 million to deal with AI safety as an issue, and £9 million was given to regulators as a whole to bid for, for innovation around AI. Those numbers were the wrong way around. If we want true regulatory innovation, a great way of doing it is pooling together pots of funding to pull the regulators together. We are seeing the FCA, the ICO and the CMA doing a great job with the Digital Regulation Cooperation Forum, but we need the lessons they are learning to be spread.

There is an additional issue: we have been consistently asking Government, I think for over seven years, to do a gaps and overlaps analysis in this space. We need to understand which new technologies are already covered by existing regulation, where there are gaps and where there are overlaps. The sort of things we are hearing from businesses—insurance companies—are, “We’re being told by the ICO not to use personal data when we’re marketing, but we’re being told by the CMA that we need to ensure we show vulnerable customers the products that are associated with vulnerable customers.” These are very confusing for businesses. It is simple, boring work in some sense, but it does not seem to be being done and it is very difficult to talk about the wider regulatory landscape unless we are having that conversation centrally.

There was the plan for—I cannot remember what it was called—a central sort of clearing office for regulation around AI, and it has been difficult to see where that has gone. I have not been so tightly involved with recent Government advice so there may have been significant changes over the last six to eight months.

John Glen: Thank you; that is really helpful.

Q104 **Dr Sandher:** Professor Wachter, how do you see AI impacting financial services and macro financial stability?

Professor Wachter: There is a wide range of ways that this is currently being implemented. UK Finance published a report last year where they said that the vast majority of deployment and efficiency gains were to do with trying to optimise and speed up existing processes rather than generating new wealth, which is important to keep in mind because at some point that field will be saturated.

There are different types of applications and tasks currently being trialled by the big banks. These include using internal chatbots that help investment bankers to give financial advice to their customers. Some banks are trying to use chatbots for customer interaction; some are using



HOUSE OF COMMONS

them to do sentiment analysis of the market to try to predict prices going forward. That is very popular at the moment.

Other things are more on the clerical side—for example, automating the process of following up on an email or drafting a contract for a client. Other very important areas where a lot of efficiency gains are being made are to do with compliance. There are a lot of reporting and transparency duties in the financial sector, and companies can use some of the tools to automate those processes. Those are just some areas where financial institutions are currently benefiting from that technology.

Q105 Dr Sandher: What about risks? The Bank of England produced a report on this in April. Basically, it is systemic risk. If AI is informing trading and investment decisions or advice from investment bankers to their clients, how concerned are you about the risks that poses?

Professor Wachter: I am very concerned about that, because the technology works by looking at the past and trying to predict the future. This will only be true if the past looks like the future, but the market is inherently unpredictable. When trying to predict what will have a major impact on the market, there are things you are not going to find in historical data. For example, financial algorithms would not have been able to predict wars or pandemics, which had a massive impact on how the market reacted. There is a lot of discussion on whether sentiment analysis is a suitable tool in terms of trying to predict how prices are going to change. Again, because the technology is not there, we do not necessarily have an up-to-date standard when it comes to sentiment analysis.

You can see this is an issue because we might rely on a system that very confidently tells us how the market is going to change, when in reality that is not something it can do. It is also a question of whether it is going to give you a competitive advantage, as everybody is getting the same advice at the same time. Again, you are going to impact the market in a way that might not be predictable.

Most importantly, who will be liable if the advice is wrong? That is something we need to be careful about because in most instances, if you make investment decisions and lose money, that is pure economic loss, which in most circumstances you cannot get redress for, so you do not get your money back. Again, we need to think very seriously about who should be liable if you lose money when making investment decisions based on faulty advice, especially because the technology does not seem to be as good as we hoped at the moment.

Q106 Dr Sandher: What about if it gets it wrong in the same direction at the same time? Before 2008, a lot of investment bankers were making exactly the same mistake: their packages of investment products reflected risk, but they did not know what was inside them. In one sense, all investment bankers were making the same mistake. Similarly, or in this particular world, how worried are you about AI amplifying those



mistakes, or each one making a similar mistake? The ECB talks about endogenous risk. Before we know it, we have a lot of balance sheets with completely the wrong situation, and then the reckoning ends up coming. How do you think Government should be encouraging us to manage these risks?

Professor Wachter: You are putting your finger on exactly the right problem. Again, when we think about how those systems understand truth—not like humans do—for AI it means, “Something is true if I have seen it often enough.” When you get those signals on the market, especially if they are wrong, the system will be more confident that this is actually what is happening. So mistakes can amplify much more quickly than they would if people are just talking to each other and making quick decisions based on that. Amplification of wrong information is spread much more quickly, and the system will be confident because it gets those signals because that is what is out there. We have to be really careful about that, especially with the financial crisis, as we already know how devastating it can be.

In the first step, regulators need to assess our current frameworks to see if this technology is causing new or similar harms that just occur or manifest in a way that the law is not prepared for. Then they need to think about smart ways of updating those standards. It does not necessarily mean we have to have new laws, but we might need to be creative in how we interpret the application of current laws. That is an exercise that still needs to be done to figure out how this technology is disrupting the current frameworks and whether it is dealing with the harms that it causes.

For example, with hallucinations, we do not regulate for truthfulness that much. In terms of standards, I am really worried that we walk into this without enough protection to make sure that deployers create systems that reduce the chances of hallucinations. Plus, there are questions of liability: who should be liable if you lose a lot of money? Those are things that are still very open, and we have to find a middle ground so the financial sector can benefit from the good that this technology can do but make sure that we do not make the market crash.

Q107 **Dr Sandher:** The big thing that worries me about the regulation, especially pre-2008, is that no one knew what was going on inside the models. No one understood the equations or the packaging. With AI, you get to a similar position where no one really understands what is going on or what is driving decisions, or indeed what they themselves are doing. So the question is: how does regulation make for that particular individual behavioural change?

At the moment, what we think of as being adequate capital ratios given leverage is actually everyone making the completely wrong mistake because AI has driven them to a certain point. All of a sudden, what you thought was—some would say—an over-prudent capital ratio is now not prudent enough because everyone is making a massive mistake and no



one realises it until of course the moment when everyone does. How do you regulate individual behaviour and actions, bearing in mind that there is nothing wrong with your balance sheet but something is wrong with the decision that human beings are making? It just feels like a different problem and a harder one for regulators to get at.

Professor Wachter: There has to be a focus on the technology itself, in addition to how we regulate the behaviour of humans. On the technical side, we need to invest in tools to understand why that system thinks that prices are going to increase or why it is a good idea to invest in this particular fund. That is important. Explainability as a field is going to be very important.

Then there is a political decision: should we be using systems in high-risk areas where we do not fully understand what they are doing? Or are there more explainable, more reliable and more robust alternatives that deliver the same result? Again, this exercise has not really been done. Is this the best course of action if we do not fully understand how this technology works and why it makes certain decisions?

The second point has to do with how we perceive AI, and how we govern people who are bewitched by it and think it is the silver bullet to all their problems, and that they are all just around the corner from being wealthy because of it. We need to discuss this much more with the people on the ground who are using those tools to make them aware that it is not necessarily solving the problem that it thinks it is. There is an exercise to be done to ask an important question. The first is, what is actually broken? What in the financial sector is broken? What needs fixing? What can be improved? Is AI the best way to fix that problem? Then you can think about a useful implementation. At the moment, it is a one-zero situation, so very binary: either you go all in AI or you do not do it at all. The harder, more prudent and nuanced way is to think: when is AI useful? What type of AI is useful? Can it deliver the promise? Or is it better to use a different type of technology to yield similar results where we have more transparency and accountability?

Chair: Professor Lawrence and Professor Andreeva both want to come in.

Professor Lawrence: I thought what Professor Wachter said was on point, but I want to highlight and draw out a couple of things. Even with £100 million, our regulators are not going to be able to do everything she said. It is actually about a change in how we approach regulation. That means more agile regulation.

The equivalent of what we are talking about is a slow-motion version of the 2010 flash crash, where the US markets lost \$1 trillion in 26 minutes. Once you have automated all these systems, and you have systemic things going on, even if they are optimal in their locality, you can get systemic risks, which are deeply problematic. The nice thing with the flash crash was that they shut down the markets. Where is the regulator's lever to shut down the market? It does not exist.



HOUSE OF COMMONS

What you need is feedback loops, as Professor Wachter was saying, where people probably do not have a particular opinion about where the technology should go, but they are very closely engaged with seeing what the effects of the technology are on the ground. That feels very different from where our regulators are at the moment. They are not so agilely integrated, but in the presence of a disruptive technology you absolutely need feedback loops to try to understand, as the situation moves, where it has gone. That involves a change of mindset in the way that we regulate and support to go through that change of mindset.

Professor Andreeva: I would like to highlight the system of internal validation that already exists in the area of credit risk, whereby all internally developed models undergo a very robust and thorough process of independent teams challenging and subjecting them to all sorts of various shocks, including macroeconomic ones. Similar systems can be adopted in any AI agents that are or will be used by financial institutions, and these agents are thoroughly tested by independent teams.

We need to create new metrics and new criteria on how to assess the success or validity of this process. To me, it is something we can utilise and build on. I totally agree that the end users should definitely be in the loop. Going back to the issue about overconfidence, this is really a very dangerous point when people become over-reliant on even a thoroughly tested AI agent. There should probably be an obligation on anyone who deploys AI in the decision-making process to explain this to the general public and to constantly work in contact with them.

Q108 **Bobby Dean:** Professor Lawrence, I was struck earlier when you talked about the data monopolies of the largest tech firms. I have a statistic here: 90% of UK banks rely on AI from Google, Microsoft and Amazon. They obviously have a huge amount of power. If companies emerge, they tend to acquire them. DeepMind is an example of that too. Is the financial services industry going to be far too much in the hands of just this handful of major tech companies?

Professor Wachter: Yes—

Bobby Dean: Sorry, that was to Professor Lawrence first. But we will come to you, Professor Wachter.

Professor Wachter: I'm sorry.

Professor Lawrence: We are all vulnerable. I do not want to name names, because I just think it is where we have got to with the current system. I was trying to think of a bit of an analogy for it. We all love to hear the first cuckoo of spring. There is an interesting reason for that: cuckoos are indicative of a vibrant, diverse ecosystem. A cuckoo cannot live unless there is wildlife and other birds for it to parasite off.

While we want to be AI makers, not takers, oddly, one of the traps we fall into is believing that somehow maximising the number of cuckoos in the



HOUSE OF COMMONS

ecosystem is going to be a good thing. We count big tech in this country and say, "Oh, it's so great that you're here. We need to do everything to make you stay." That is misunderstanding the arrow of causality. Those companies want to be here because they build on existing businesses in the way a cuckoo builds on a robin's or a reed warbler's nest. If we increase the amount, they will destroy, and we will go through an ecosystem collapse, which is the best trajectory we are talking about at the moment.

The funny thing about that is that where big tech companies claim they are the ones doing the research—which is palpably untrue—it is not just universities that suffer: it is the FTSE 100 and financial services companies too. It is not in the cuckoo's interest to lose the robins and the reed warblers; it is in the cuckoo's interest that the ecosystem exists—but, of course, it is in the short-term interest of the cuckoo. That is the real problem we face.

I am going to avoid the obvious joke about cloud computing and cuckoos, but the real issue we have is that people misunderstand the presence and the excitement of large tech in an ecosystem as something that we should be directly supporting, as if we should breed cuckoos. We should value and love the cuckoos that are here but actually worry about the robin, the reed warblers and everything else, which is the diverse ecosystem, including large financial services and FTSE 100 companies that are also vulnerable, as well as our teachers, nurses, hospitals and everything else. It is a severe danger.

Bobby Dean: I am not sure I have fully kept up with all the cuckooing there.

Chair: Last week it was PuFIns.

Q109 **Bobby Dean:** The thing I am getting at is that it might feel like a benevolent power in the market now, but standard economics tells us we should be very distrustful of any individual companies that have this amount of power. You have seen their appetite for acquiring competitors. In recent times, we have seen tech companies taken over by people who are very politically active and have their own agendas that may not match that of our country or the industries that are operating there. We also have a quote from Nikhil Rathi, who said that the big tech companies could buy up pretty much any of our UK financial firms if they wanted to. Is it a specific threat to the UK financial service industries if you have a handful of large American companies to depend on?

Professor Lawrence: It is a specific threat across the UK, including financial services—which ironically is the area where we are traditionally strongest economically—and it is a massive one. Companies go through an interesting transition. Traditionally, they protect their market through what are called APIs, which are specific software interfaces that they control. Previously, a lot of the work has been around trying to get them



to standardise APIs so we can have, for example, the equivalent of containerisation for financial or medical systems.

The really interesting thing about this technology is that the API is English, so you do not need to standardise it. Computers can now talk to each other directly. That is effectively what people mean by agentic AI. I can take Claude, the anthropic model, and replace it in moments with ChatGPT or with the Google model. So instead of having a lockdown market on search that they have had before, they are very exposed in this area. You can see them actively manoeuvring, as a Google internal demo said, to protect their moat. One of the ways they do that, and their most important and useful approach, is to exploit existing market dominance.

In terms of our regulators, particularly the CMA and the Digital Markets Unit, which have done an enormous amount of thinking in that regard, we have to think about how to correctly regulate this incredibly useful service. Let us also be clear: what they provide is incredibly useful. It is shocking that no one speaks about how narrowly the Digital Markets, Competition and Consumer Act managed to get through at the end of the last Parliament—it was given Royal Assent on 24 May. It gives the Competition Markets Authority extraordinary powers to deal with asymmetries that arise from data after the assessment of strategic market status.

Rather than seeing this as an opportunity to create space for UK companies and ensure that large companies are regulated in the way that we would like, and so they do not have increasing power over our financial services sector, it has been disappointing that there has been a large amount of pushback against the Competition and Markets Authority from the current Government. This is a shame because there is an enormous opportunity within the UK.

I do not know whether people are aware of strategic market status. After a nine-month investigation, the Digital Markets Unit can put companies under this status and place them under very enhanced regulatory provision with fines of up to 10% of global revenue. These are extraordinary powers that originate back to the May Government, and we are incredibly lucky to have them. We should not leverage them quickly, but we should be considering how to deploy them across the sector to ensure what you are saying does not happen.

Q110 Bobby Dean: Professor Wachter, do you want to build on that point? Given that the Competition and Markets Authority has just been taken over by somebody who was previously very high up in Amazon, I am not encouraged that this kind of action is going to be taken. Would you agree or disagree?

Professor Wachter: I fully agree with Professor Lawrence's statement. The only thing I would add is that over in the European Union and in the member states, there are serious movements to think about digital



HOUSE OF COMMONS

sovereignty, exactly for the reasons that were just mentioned. That means that on a European and a member state level, they are actively thinking about how to decouple themselves from the market monopolies that some US companies have.

There are some interesting strategies here. For example, the German Government have said they are going to have an increased focus on trying to invest in and use tech that is German. We see similar things on the European level that are very interesting. In the last couple of years, there have been a couple of regulatory frameworks that are supposed to induce new healthy competition. We have the Digital Markets Act, the Data Act and the Data Governance Act, which are trying to rebalance the market powers, if you will. It will be easier for start-ups and SMEs to get data access from companies or the public sector, on top of which to build their own businesses. There are strategies to make sure there is enough interoperability between systems, so that it is easier for customers to switch to a different operator if they wish. There are a couple of very interesting, clever and creative ways of trying to induce healthy competition to make sure that we have more offerings.

I would find it interesting to see whether similar strategies are possible in the UK, where data streams could be made available to boost businesses here that could then, for example, offer an alternative to the products that we see in the US.

Q111 Rachel Blake: I want to pick up on the themes of accountability, empowerment and the transparency of models, and the markets and the trading aspect we started off with. The Governor of the Bank of England has talked about not knowing what is happening inside AI models. We know from the artificial intelligence in UK financial services survey that 46% of respondent firms had only a partial understanding. Professor Wachter, what is standing in the way of the type of crash that we had in 2008, which was driven by a lack of understanding of what was happening in certain transactions?

Professor Wachter: I think you already answered it. Unfortunately, we are at a great risk because there is very little technical understanding. It is not even a matter of blaming the people who use it, because not even those who built the system fully understand what goes on inside the black box. It is a real issue that we embed and trust systems that sound convincing and spit out plausible-sounding text that has nothing to do with truth. We have a technology bias, so we tend to believe in technology much more than we believe in humans.

As soon as we see language, we are tempted to think that we are talking to another being that understands our intents, motivations and feelings, and gives us something back that is reflected upon. We fall into this trap because we anthropomorphise technology. It is even trickier because those systems are created to make us think that they speak the truth. This is an actual process that happens when creating those systems. It is



called reinforcement learning with human feedback, where people are hired by tech companies to look at—

Q112 **Rachel Blake:** Sorry, but I just wanted to follow up. If there is an acceptance that there is a significant risk, what is standing in the way of your proposals around counterfactual explanation statements being some kind of requirement in order to add to the transparency of these models?

Professor Wachter: It would definitely help, because it would tell you how a specific decision was made in a certain instance. For example, if you were taking out a loan and you did not get one, a counterfactual explanation would tell you, "You didn't get the loan because your income was £40,000; if it had been £50,000, we would have offered you the loan." So it tells you something about the individual decision and how it was made. That helps the customer and it helps the bank, because you know that income was the issue here. So that is really helpful.

Rachel Blake: Google, IBM, Accenture and Microsoft have adopted these systems.

Professor Wachter: Among many others, yes.

Rachel Blake: So what is holding back other organisations from adopting them?

Professor Wachter: That is a good question. I am not sure.

Q113 **Rachel Blake:** Professor Andreeva, what do you think is holding back other companies from adopting these systems?

Professor Andreeva: That depends on the area of application. In my area of credit risk, I am not familiar with a huge use of generative AI, but I would advocate that when this adoption happens, there will be a thorough validation and testing that I mentioned. I would not be that prescriptive in recommending any particular measure that everyone should be using, but there definitely should be a range of different measures and a lot of brainstorming in terms of what is best or state-of-the-art in measuring the adequacy of AI agents and models. But it can be a legal requirement for all financial institutions to have due diligence in place.

Rachel Blake: On consumer and on trading?

Professor Andreeva: On consumer and on trading and credit risk—any area where AI agents are or potentially can be employed in the future.

Professor Lawrence: It is a really interesting topic that cuts to the heart of the challenge we are facing. If we think in terms of consumers, first, we want ecosystems where providers are incentivised to do good things for their customers. That is why some companies you mentioned are, for example, adopting Professor Wachter's counterfactual explanation approach, because it is in their consumer-facing interest. One of the challenges we face in particular in financial services is that we get



HOUSE OF COMMONS

economies of scale by separating from the consumer. That is what companies do by removing and trading things at scale. That is trading. Then we cannot expect to see the same kind of direct consumer signal that encourages these types of inputs.

When we look at the history of how we have tended to deal with this issue, our tool was statistical monitoring of what was going on to try to identify problems as and where they were happening. I have a fear of something I think of as the big data paradox. I have talked about this for about a decade, but we are seeing it play out now at a time when, oddly, in society we are getting more and more data, but we are understanding less. It is super interesting. I do not have the full explanation for what is going on, but you see the Office for National Statistics struggling to produce the type of classical statistics we want.

During the covid pandemic, I gave advice to the Government. With colleagues from Imperial and Warwick, we were able to monitor minute by minute the GDP of Spain, because there was a deal between BBVA, one of the largest banks, and these economists to look at payments data as it was moving through the system. We could see the impact of non-pharmaceutical interventions in the Spanish economy on localities in terms of what was happening in real time. There are some great papers on it.

We could not do it for the UK because we could never get around to accessing the right type of data, even though at the time Mastercard was actually very willing. But after that, we recommended there should be a statutory obligation on businesses to provide some of this data as a matter of course, including anonymised data on the movement of people and payments. Governments should not have to stump up £9 million a year for anonymized data from our own citizens; it should be the case that this data is managed in a way that allows us to more finely monitor these ecosystems.

I have much respect for Professor Wachter's work and what Professor Andreeva was saying on the consumer-facing side, but we need better statistical monitoring techniques to understand non-consumer-facing issues when problems occur in real time.

Q114 Rachel Blake: If we had better statistical monitoring techniques, do you think would we be able to read the black box models more effectively ourselves?

Professor Lawrence: The notion of the black box model is a little overdone, because as soon as you build a system and attach all these things together, as in the flash crash of 2010, you have created a black box even if you understand the components. Systemic risk comes from the interaction of these components in society. You can prove whatever you like. I could prove that a driverless car is never going to hit anyone. What does that mean? That means a 15-year-old can steal the car by putting a cardboard box around it, walking off and towing the car with a



string, because it is not going to hit the cardboard box. These are the unforeseen things that happen when you hit society. It is very difficult to prove.

The Chair looks confused. Imagine I can prove a driverless car is never going to hit anything. All I have to do is put a box around it, and as I walk away with the box, the car will be forced to move with the box. It is the sort of thing a 12-year-old will get up to as soon as they realise that this is the case. One of the beautiful things about humans is how they react in that circumstance. They will just drive through the box or get out and shout at the kid. The idea that we can verify these systems to 100% is wrong. As soon as you deploy these things in society, unforeseen corner cases occur. That is where the problem is.

Chair: If it is real-world.

Professor Andreeva: I wanted to quickly mention the power of financial transactions data and that it is already partially available to the Government and researchers. The University of Edinburgh holds a vast dataset from one of the biggest banks. I totally agree that analysing financial data can open huge advantages in trying to see macroeconomic shocks developing in real time, not waiting for survey data or other sources of information.

Q115 **Rachel Blake:** Professor Lawrence, where do you see the main gaps of atomic humans at the moment?

Professor Lawrence: In terms of our ability to do things in practice?

Rachel Blake: Yes. I am pretty sure we are the strength, but where do we need more of us in order to mitigate some of the risks Professor Wachter has been talking about?

Professor Lawrence: That is really key. As I said earlier, all the risks we are talking about at the moment are of the digitisation we have already been through. We are already experiencing them, and they are already live. The choice space in front of us is one where we reinsert the human into these systems so that we understand early that the Horizon system is making significant errors, particularly for rural post offices, and we can intervene to support rather than having to wait downstream for 15 years when it has become obvious. It is not clear how that is going to look, but one of the extraordinary realisations you have is that this technology, while it is not AGI, is so transformative that in the future it will just be seen as totally anachronistic. It will utterly change the way we do everything.

It is very easy for a disruptive start-up company to start from scratch, but when we are talking about Government, healthcare, education or even established businesses like financial services, they have to understand how to assimilate these technologies to get the advantages, while not losing the vital role that they play in society.



HOUSE OF COMMONS

I do not know the answer. The first step is being honest about the fact that we do not know the answer. At the moment in Cambridge, we are focusing a lot on public dialogues, talking directly to citizens and integrating as many people as possible, because it is interesting how much better their answers are than a lot of those we hear from big tech companies.

Q116 Yuan Yang: Professor Lawrence, I want to come back on your point about the ONS data. This Committee has heard from the ONS and interrogated the ONS about the problems with the labour force survey and other national statistics. There is currently a Cabinet Office investigation into the ONS. Do you have any advice for how the ONS can make more use of the AI data that you just mentioned?

Professor Lawrence: I do not know the details of everything that has gone wrong. I have to say I worked near Sir Ian Diamond during the pandemic, and I wondered where, when we urgently required data, everyone went. They went to the ONS. They trusted the ONS to have and store their data.

There are obviously challenges for an organisation that is expected to focus on what we think of as gold standard statistics, and we must not allow those statistics to be lost, but simultaneously we must try to innovate. One of the things I have been excited about is the data science campus in South Wales, where, for example, they are coming up with alternative inflation measures driven by modern data science techniques.

One of the things I was quite horrified about, and mentioned in a keynote talk I gave a few years ago, was the idea that we would be able to get rid of these surveys and replace them with new forms of data-science-driven surveying, which I think is what has driven the reduction in their budget. It is so important that we have the gold standard to compare against for these new statistics—we need both. One of the things we have seen is too much overemphasis on what this brave new world is going to be like, without realising that we need the handrails of the old world to hold on to as we introduce these new statistics and understand what they mean for our economy.

In the long term, of course, they will be much better for allowing us to make agile performances, but again, I do not know whether this story is about culture or leadership or whatever else. My experience is that we vitally need them, they are the goose that lays the golden eggs, and we must do as much as we can to ensure that those eggs are being laid while we are going through these transitions.

Q117 Yuan Yang: Thank you, Professor. I will return to our main theme of AI and financial services, and in particular on bias in AI. A lot of writing about AI risks focuses on medium or long-term risks that have not yet occurred, but Professor Andreeva and Professor Wachter have argued that the growing use of AI in decision making increases the risks of discrimination against certain groups, and that in practice this is



happening already. Professor Wachter, could you give some examples?

Professor Wachter: It is really helpful to think about how AI decision making works: it is always looking at past decisions and trying to predict future decisions based on them. So whenever you are in a sector that has seen discrimination, be it healthcare, employment, the criminal justice system or the financial services system, as soon as you are using historical data on who got a job, was sent to prison, got a loan, or was admitted to university, AI will pick up the patterns of those decisions and transport them into the future. What AI cannot do is to say, "Stop. Are you aware that you are not giving out loans to women at the same rates that you are giving them out to men?" This critical thinking is not something that AI can do. It repeats the same patterns that it has seen before and transports them into the future.

For the last 10 to 15 years, many people have filled libraries where we can see systemic discrimination in employment as soon as AI is used. We see underserved groups in healthcare because of bias in the data. We see that bias in the criminal justice system, especially for people of colour, is amplified because of the use of AI. We also see it in the financial sector: redlining has become more prevalent because AI is able to indirectly infer ethnicity, for example. That is something that has negative consequences for society. In my opinion, we need to use the reverse of the burden of proof: unless you can prove the opposite, I will have to assume that your data has a bias problem. That is a much more helpful way to think about it.

Q118 **Yuan Yang:** Professor Andreeva, Professor Wachter mentioned redlining; have you come across any other examples of discrimination from the use of AI in financial services?

Professor Andreeva: It is a much more complicated problem than the majority of people realise. I totally agree that any model built on historic data—be it traditional statistical models or AI—will reflect the inequalities and discrimination embedded in that data. We live in an unequal society and AI, due to its power, reflects these inequalities better than statistical models.

You can train statistical models and AI models to give certain outcomes. You can set it as a criterion, for example, that women and men should receive a certain proportion of loans, or an equal proportion of loans. However, there are real problems with this approach. First, equality law forbids the use of personal characteristics, which means you cannot ask AI to use protected characteristics for monitoring because it would be a violation of equality law. Even if it is allowed, there is no data to test the outcomes for equality, because bankers got so frightened of being accused of discrimination that they stopped collecting sensitive data. We have age and gender, which can be inferred from marketing information, but that is it. Other protected characteristics are not collected.



There is also a lack of agreement in terms of what the fair outcome should be. I gave you one example of an equal proportion of men and women being accepted for loans. That is just one fairness metric; there are 20 other metrics. Everyone is confused in this space. We need clarity and guidance in terms of what can be done and what cannot be done. Equality legislation is making things a little more complicated.

Q119 Yuan Yang: Professor Andreeva, you mentioned that some banks have been so frightened by the accusation of discrimination that they have stopped collecting data on protected characteristics, which means they cannot effectively monitor whether their outcomes are discriminatory. Do you see any role for the regulator in monitoring these kinds of outcomes in financial services?

Professor Andreeva: Absolutely. As I mentioned, it would certainly be helpful to give some clarification on what is and is not acceptable in the space of protected characteristics. However, the problem is deeper than that. I represent the Credit Research Centre and a fair credit working group that comprises academics and practitioners. Following my written evidence submission, I received some feedback saying that even if bankers are allowed to collect protected characteristics, they will not be brave enough to ask for these characteristics from their customers because they fear it would appear unethical. Maybe the way forward is to create a data depository or golden sample that can be used to test the quality of outcomes.

Q120 Yuan Yang: On that theme, have any members of the panel had discussions with the FCA or other financial regulators on tackling discrimination?

Professor Wachter: I had a discussion some time ago about this; it was more focused on equality law back then, for two main reasons. First was the idea that new groups that are not protected under non-discrimination law currently might be hurt. We have a finite number of people who are currently protected under the law, but AI might start discriminating against other groups that are not protected under the law yet. A toy example would be that AI finds a correlation between credit worthiness and dog ownership and assumes that somebody who does not have a dog should not get a loan. Obviously, a dog owner is not somebody who is protected under the law, but for reasons that are not immediately apparent, AI could see this particular group as relevant.

To give another toy example, if you think about tenancy rights in London, most landlords will not allow people to have pets in their flats, so if you have a dog in London it is more likely that you own your own property. Suddenly, a proxy for wealth has been created that you do not know about.

This is quite an interesting problem to deal with: that we might have new proxies or new groups that we might discriminate against and that the



law did not anticipate but that are none the less relevant for financial decision making. That is one of the things we have discussed.

- Q121 **Yuan Yang:** Very briefly on this point, Professor Wachter, you have created a test—the conditional demographic disparity test—to check for discrimination. It has been incorporated into, for example, Amazon’s toolkits. Do you see a role for regulators in encouraging or even requiring this kind of bias detection tool?

Professor Wachter: Absolutely; bias testing should be mandatory. IBM and Amazon have both implemented this test now. It is also wonderful to hear that an NGO in the Netherlands used this test one or two months ago to uncover systemic discrimination when giving out financial aid for people wanting to attend university. The Minister for Education apologised for that, and they are now changing the processes, so this test can actually help to uncover discrimination.

The test is also really important because, as we have already heard, there are so many bias tests out there that it is really hard to know which one to choose. In the past, we looked at the most common bias tests that are currently being used, and we found that two-thirds of them are not legally compliant with UK and European non-discrimination laws. The reason is that most of them were developed in the US, where the equality laws are fundamentally different. Again, this is important for people to know.

You might want to do the right thing, you might be a financial regulator or a company that wants to test if its systems are biased, but you do not know that you are maybe clashing with the law because you are using a test that does not adhere to the rules. It is really important to give clear guidance on which test to use and when, and that this bias test can definitely service that.

- Q122 **John Glen:** Turning to the insurance industry, clearly the amount of data points that exist for different individuals allows companies to form a different profile and understanding of risk, which has quite significant implications in terms of the loss of pooling, the anatomisation of a market and the offering of different positions. Do any of you have anything to say about the current awareness of what data is being used and what the risks are? For me, it seems quite fundamental. Fintechs—we have a lot of them in the UK—could track a lot of data and offer an undercut, essentially, by not having to take the pooling. That is my instinctive reaction to what could happen. It would be great to get your views on what consumer awareness there is around this. What is actually happening, and what do you foresee could evolve in the coming weeks, months and years? Professor Lawrence, would you like to start?

Professor Lawrence: I will make some broad comments, because this is something we have seen coming for a while. I cannot say I have tracked the latest developments, but my understanding is that insurance originated—in the UK, at any rate—with Lloyd’s Coffee House, where



people insured their ships. You can see how that made a lot of sense because it was hedging, in a really interesting way, for uncertainties that are difficult to deal with and manage.

Insurance-obsessed people have done a load of analysis on Laplace, but they never had to face this question of what it means when the insurance relates to the insuring of yourself, and you are paying for it. There is clearly some extreme version of this where it makes no sense. Between those two extremes, there is a point at which society becomes interested, maybe to a different extent, according to whether it is life insurance, some forms of health insurance, or pet insurance, perhaps.

I do not know the right answers, but it is clear that there is something quite odd going on when insurance has shifted from being a way of spreading risk to a way of spreading fear, where you end up telling people, "You have to be frightened of this: it will do this to you. You need to insure against it." If that is the way we are going with insurance, it is not a great place. But I am afraid I am not deeply expert on what is happening, other than to have seen it coming from about 10 years ago. My colleagues are likely to be more expert.

John Glen: Professor Andreeva, do you have anything to add?

Professor Andreeva: In terms of the data used by insurance companies, this is mostly historic data from previous accidents and insurance claims, including the costs and any related, correlated characteristics that are connected and that are legally allowed.

In terms of what consumers know about the data being used, it is difficult to answer. We probably need a consumer survey to answer this question, but requiring insurance companies to fully disclose exactly what information they use on consumers may undermine the confidentiality and commercial sensitivity of their models and may open the risk of abusing the system.

Q123 **John Glen:** Is there not a significant risk that if we do not have more intervention in this area, we will actually have market providers diminishing as their data confidence levels over risk levels for some cohorts in society become so high? That imperative around privacy actually gives them the right not to offer a service in order to maintain a more profitable market segment.

Professor Andreeva: That is a danger. However, there are already regulatory requirements that protect against it, in particular the FCA consumer duty, which makes it clear that customer needs should be the priority when developing and offering different products.

Having said that, financial inclusion remains a big problem, and I can totally relate to your remark that there are pockets of underserved customers. Data may be an issue—for example, some segments may simply not have the data required for credit risk assessment or insurance



assessment. That is why an exploration of alternative data is the way to go. There are experiments happening in this space.

Professor Lawrence: I would like to follow up on that, because I thought what you were talking about was a very good example of the challenges we face around monitoring and enforcement. We can talk until the cows come home about what regulators should do and what we would like people to do, but we have limited budgets. We need more agile approaches. We need to understand more rapidly when something is going wrong and have the levers to intervene, and I do not think we have enough conversations about what that looks like. Those conversations obviously need to involve regulators. Of course, we are having that conversation here, but it is not dominating our psyche across a number of sectors.

Professor Wachter: To echo the statement of Professor Lawrence, in the most basic sense, insurance is a bet: the insurer bets that I will stay healthy and I bet that I will get sick, and for every month that I am not sick, I basically have to pay a fee because I lost a bet.

With the new tools that we have right now, one side starts to cheat because suddenly I am becoming completely transparent, the insurer knows everything about me and I know nothing about the insurance company, so that it is not a fair bet any more. It also goes completely against the original idea of insurance, which is risk pooling for an uncertain future. Then, a system is created that sorts out people who might get sick, keeps the people who are going to stay healthy, and maybe even charges them higher rates because you find out that they are in strong financial health. That is really not what the business model is about.

Q124 **John Glen:** Just to clarify, Professor Wachter, do you see the regulators doing enough at the moment to deal with the evolving risk in terms of the use of this often protected data?

Professor Wachter: There is room for improvement, especially because a lot of these companies are starting to use untraditional data, where we do not necessarily know how it links to protected characteristics. We know about the example of redlining because that is related to postcodes, but now they will use your training data—your workout data from your health app, for example—or monitor other untraditional data sources. We do not necessarily know what is being inferred from it, so we do not have the same alarm bells going off. We really need to think about how we regulate untraditional data uses when it comes to insurance decisions.

Professor Lawrence: Professor Wachter put that beautifully. To add to those who are not in on that asymmetry in the bet, it is the Government and regulators themselves. How can they monitor the asymmetry that is occurring when they do not have access to the data either? We are in an extraordinary modern position. People used to worry about the situation



where the Government controlled us through our personal data, but the Government do not have access to that data.

John Glen: Regulators and Governments are always very reticent to put imperatives to the tech companies for fear of over-regulation.

Professor Lawrence: But there are imperatives. My understanding from a recent talk by my esteemed colleague Dame Diane Coyle is that they are not fulfilling their current obligations.

John Glen: Yes, quite.

Professor Lawrence: It is not that we are reticent; they are reticent, and we have very little means of enforcement.

Q125 **Dame Harriett Baldwin:** It is wonderful to have such expertise with us this morning. What our constituents will be concerned about in particular, when they think about artificial intelligence in financial services, is how secure their data is, and in particular how much more it would increase their vulnerability in terms of crypto and cyber-security if quantum computers are using this data and are able to penetrate the encryption in the financial services sector. Professor Wachter, you are nodding the most, so I will come to you first. What are your views on that, and how you might reassure our constituents?

Professor Wachter: Cyber-security and fraud are definitely a risk on multiple fronts. On the one hand, malicious actors can use it to launch cyber-attacks. They can also use it to defraud people. You just need a very short portion of someone's voice and you can create a convincing phone call to extract money from somebody. Many of you will get spam emails in the same amount that I do, which are clearly phishing attempts; however, generative AI is much more sophisticated, to the extent that we might fall into a trap. There is definitely the risk of financial ramifications, but on the other side, there are ways of using those tools to combat these issues.

Again, it is very important to be cautious; fraud detection is a complicated and nuanced game. Unfortunately, there are problematic examples in Sweden, the Netherlands and the UK, where fraud-detection software was used to assume that somebody did something wrong and it was completely not true. It is a very important area that we have to take seriously. Yes, cyber-risks and fraud are going to increase. We can use some tools to fight back, but they have to be properly tested. This idea of having a perfect algorithm that will find out all the threats is not going to be possible.

Cyber-fraud is always going to be a cat and mouse game, and that is just how it is going to be probably forever, because as soon as you build a strong system, you will find somebody who will be able to break it. You will find somebody who can repair it, but the next person will break it again. This just means that we have to be very watchful and see what new techniques are currently being used to commit fraud or cyber-



HOUSE OF COMMONS

attacks, and be ready for them, but we must also acknowledge that systems can be wrong in detecting fraud and malicious actions. We must make sure that we do not hurt people by assuming that they did something wrong.

Q126 Dame Harriett Baldwin: Professor Lawrence, I assume you recognise that picture. Would you be able to comment on the payment architecture itself and how vulnerable it is to the power that artificial intelligence might be able to give to the malign actors in this space? Is the Bank of England doing enough to protect itself from vulnerabilities that might arise because of the power of artificial intelligence?

Professor Lawrence: I do not have enough knowledge of the precise payment architecture, but I can perhaps step back and say a few things around the context of what we might expect and give a few numbers that I always find interesting.

The last time I looked, the budget of the National Cyber Security Centre was £2 billion a year, but they are struggling to deal with the cyber-attacks we already face. On the basis that more AI is going to mean more software and a very different type of software, we can expect more vulnerabilities. Payment architecture will tend to be protected from that because people will be more careful; it is a similar situation with medical systems. That is why regulation is important, but we can expect more vulnerabilities.

The other problem we can expect, unfortunately, as Professor Wachter said, is that this is a cat and mouse conversation, and in this case, the cat is far more agile. People who are engaging in cyber-attacks will be able to adopt these tools far more rapidly than larger, inertia-bound organisations will be able to adopt the tools that prevent attacks.

As we were just saying, this is ironically an area where the insurance industry could help. There is an emerging cyber-insurance industry, and we need it to be healthy, because one of the big challenges, by my understanding, is that companies are not willing to share when they have been under cyber-attack. We can all be extremely grateful to the British Library for sharing what they went through. Many organisations do not; they try to keep it quiet, which is leading to a wider problem where we do not have a good enough service industry in terms of supporting small, medium and large enterprises and protecting their cyber technology. Without a change in that, I would not be able to confidently predict the future.

As I have said, confident predictions are always wrong, but you could imagine a world where in a decade it is costing us £20 billion a year to run the National Cyber Security Centre, and it is still not covering everything, which is a real worry.

Q127 Dame Harriett Baldwin: Just as we have a Pool Re for terrorist risk, do we need a Pool Re for cyber-security risk across the financial sector?



Professor Lawrence: That may have been the way we were thinking about it. I do not have the statistics to hand, but my sense is that it is far more common than that. It is not unusual for an organisation to be under cyber-attack. It is far more common than a fire, flood or any of these issues that we are insuring against now. Of course, the benefit of insurance is that people, where they are able, can change their risky behaviour. An interesting insurance area is where people have devices in their cars to monitor their driving. One can say whether one likes that or not, but certainly young people who drive carefully are getting better insurance deals. In this case, we would prefer to imagine a world where we can run those things commercially. Why would we want to regulate it? It is very difficult to intervene on regulation where incentives are aligned, so that organisations with better cyber-practices are paying low insurance premiums.

The problem we actually face is that we do not have an expert enough insurance industry to understand what the cyber risk is in a given business, and the scale of this problem is already much larger than we can perhaps accommodate. But it is a vital threat to security at all levels, from personal risk to trust in commerce: it is the new security front. Britain is an island, which gave it some advantage for defence in the past, but the internet is not an island. Everything is connected, and most of the attacks we are under at the moment are either geopolitical or bad actors attacking our existing systems. One can only feel that it will get worse before it gets better.

Dame Harriett Baldwin: I am not feeling very reassured.

Professor Andreeva: I totally agree with the other witnesses: it is an ongoing battle. AI can be used by hackers, but equally it can be used to prevent hacking attacks. Cyber-insurance is an interesting proposition because it can be an incentive—a motivation to collect the data on incidents. The lack of disclosure or the fear of disclosing these attacks is a major roadblock to actually creating defences.

Q128 **Chair:** Professor Wachter, you have said—we touched a little on this—that the EU legislation on AI has, “An over-reliance on self-regulation, self-certification” and “weak oversight and investigatory mechanisms”. Do you think the same is true of UK legislation and regulation on AI?

Professor Wachter: It is hard to say at the moment because we are at the beginning of trying to figure out how—if at all—we are going to regulate AI in the UK. The most important thing is that there might be a couple of lessons learned from what the European Union has done that we could make sure we also do. One of the things I criticise, and I stand by it, is that the vast majority of risk assessment is done internally before systems go to market, so there is an increased risk that those systems will actually harm people. There is not enough oversight there.

Q129 **Chair:** You have been very clear that big tech has watered down the EU legislation. What got watered down that you would have wanted to see in



there?

Professor Wachter: There are three things, the first of which has to do with liability. As I said, intangible harms are not compensable under the directives that the European Union established. That was definitely a lobbying effect, and intangible harms, such as privacy and discrimination harms, pure financial loss and pure economic loss, are the most common. Those are not the ones you get compensation for, and that is a problem.

Secondly, classifying only a very limited number of financial services as high risk was a missed opportunity. In the European Union, only credit, health and life insurance are seen as high risk, but other things like wealth management, institutional private investment, and financial trading are not seen as high risk, even though the ramifications for the whole economy could be great if something went wrong in those instances.

Thirdly, there is nothing in the EU AI Act that requires developers to create systems that curb the chance of hallucinations, nor is there any mechanism for liability. If the AI hallucinates and says something that is untrue and you act on its advice, the companies would not be liable for that or have an obligation to make sure that system hallucinations decrease going forward. Those are the three things that are most relevant for the financial sector that we could take the opportunity to not repeat again.

Q130 **Chair:** We have an opportunity to do that, of course, with the Artificial Intelligence (Regulation) Bill, which was introduced in this Parliament in March. What do you think about that Bill?

Professor Wachter: I love it. It is a fantastic suggestion. There is something that I find very reassuring about having an AI authority that does exactly what I do for a living, which is going through regulation to figure out if it is still fit for purpose, and what to adapt and how to adapt it.

It is also really important to have a co-ordinating function, because not only does AI not believe in boundaries or geographical limits; it does not care about different sectors. This is a data protection issue, an IP issue, a market issue and an employment issue. It is going to be really important to have an authority that can co-ordinate efforts to have a system that creates AI that makes everybody better off rather than worse off. That is really great. I love that there are efforts in there to make systems less biased, and it is fantastic that there are transparency requirements and labelling duties. The whole IP issue is also taken up, which is an ongoing battle that we need to think about. So yes, this is a fantastic first step in the right direction, and I am a big fan of the framework.

Chair: It is great to finish on a positive note, because we have heard today of the huge risks of bias and discrimination. There is a huge opportunity, but the risks need to be balanced. Thank you very much to



HOUSE OF COMMONS

our witnesses for their time. Thank you to our colleagues at Hansard and to our colleagues at Bowtie for sorting out the virtual and other broadcasting arrangements. The transcript of this session will be available on the website uncorrected in the next couple of days.