



Communications and Digital Committee

Corrected oral evidence: AI and copyright

Tuesday 9 December 2025

2.35 pm

[Watch the meeting](#)

Members present: Baroness Keeley (The Chair); Viscount Colville of Culross; Baroness Elliott of Whitburn Bay; Baroness Fleet; Baroness Healy of Primrose Hill; Lord Holmes of Richmond; Lord Knight of Weymouth; Lord McNally; Lord Storey; Baroness Wheatcroft.

Evidence Session No. 4

Heard in Public

Questions 56 - 78

Witnesses

I: Vinous Ali, Deputy Executive Director, Startup Coalition; Matthew Sinclair, Senior Director, Computer and Communications Industry Association (CCIA); Antony Walker, Deputy Chief Executive Officer, techUK.

USE OF THE TRANSCRIPT

This is a corrected transcript of evidence taken in public and webcast on www.parliamentlive.tv.

Examination of witnesses

Vinous Ali, Matthew Sinclair and Antony Walker.

Q56 The Chair: Good afternoon and welcome to this meeting of the Communications and Digital Committee. I am Baroness Barbara Keeley, the Chair. Today we are continuing our inquiry into AI and copyright, and we will be hearing from two panels of witnesses. I am pleased to welcome our first panel, who are representing tech industry bodies. Both of today's sessions will be broadcast live and transcripts will be taken. Our witnesses will have the opportunity to make corrections to the transcript when necessary.

I shall start with the first question to open the discussion. At the heart of what we are trying to get to is the UK copyright framework and how it can sit effectively alongside the decisions made by AI developers. I ask each of you: do you think it is proper that copyright holders are reimbursed for their work?

Matthew Sinclair: I do not think that learning was ever what copyright was intended to police. We all learn from publicly available information. AI models do the same and I do not think that should trigger reimbursements. However, absolutely, when copyright is used in the outputs of models or when distinctive datasets are creating distinctive value then there is space for licensing deals, and we are seeing those deals announced on a regular basis—to be fair, mostly where most of the training is taking place, in the United States, which I think in part reflects its more supportive copyright regime. Fundamentally, we should draw a distinction between what copyright is intended to police and learning, which it was not intended to do, and ensure that we broadly match the kind of policy outcome that we have seen that recognises that in the US, the EU, Japan and other major economies.

Q57 The Chair: I asked you about our UK copyright framework, which is not fair use. It is a different set-up, isn't it?

Matthew Sinclair: Sorry, just to be clear, while it is not fair use, fair use is a means of getting to an outcome of copyright. It does not change the fact that copyright in the UK, as well as in other countries, has never been triggered by people learning from information. We all learn, and that contributes to our ability to generate all kinds of value, but it has never triggered a copyright claim. We would see AI as another version of that.

Vinous Ali: For me, copyright—despite working in this policy space for a while now, I do not claim to be an expert in the intricacies, but this is my understanding—protects the expression of ideas in their fixed formats, rather than the very facts themselves. Similarly to Matthew, I would be very comfortable with looking at the outputs end, but training AI models is about the permission to learn on a machine-level scale and that is therefore distinct from the copyright regime as it is currently and as I understand it to be pursued.

Antony Walker: Copyright is a fundamentally important mechanism for enabling innovation to take place. Indeed, techUK has almost 1,200 member companies, many of which are significant rights holders in relation to entertainment or sporting content, significant scientific journals and so on, so copyright is fundamentally important. The question is the role that copyright plays in relation to AI development and AI training. That is where the US approach and what happens in the US jurisdiction is significant because the models are largely being trained in the US under the US fair-use approach.

Q58 **The Chair:** I understand that, but we are concerned with the UK copyright framework, which is different. We should keep focusing on that.

I will move on. We have heard from witnesses that there is a huge lack of volition on the part of AI companies to license training data. Indeed, it has been said to us that there is a critical mass of AI companies that have just not bothered to license training data. One of our witnesses on the next panel is using training data that is licensed, and we know from other witnesses that there is a serious and growing market for licensing data; news, publishing and photography are all in that licensing market. As I mentioned, at the heart of our inquiry we are looking at how our copyright framework in the UK can be effective alongside the decisions made by AI developers on how to develop, train and deploy their models. That is what we are working through—how can these two things work together? We are trying to do it here and the Government are trying to do it through the technical working groups.

Vinous Ali: I am happy to come in on the specific question of licensing. From our perspective, we have seen licensing deals emerge. They tend to be between large publishers, which have very deep pockets, and some of the large model providers. As Matthew alluded to, most of those deals are happening in the US because of the broader permissive regime there.

I completely appreciate that the committee's focus is on the market here in the UK but, given that we are operating in a far broader global race and we are seeing global fragmentation, there is a question around how we ensure that the UK remains competitive in that race.

For that reason, we feel that there needs to be a more permissive regime here—that is, a broad TDM exemption—that allows training to take place here in the UK and does not sit against or contradict a licensing regime that works. However, the way in which things are currently drawn means that we are seeing licensing deals happen for undisclosed sums of money; UK AI start-ups have absolutely no possibility of engaging with them because, frankly, their runway and pockets just are not as deep.

Matthew Sinclair: I would just distinguish something a little because some of these models are trained on, broadly speaking, the internet—that is, billions, if not trillions, of websites—and the value of any individual piece of content in that is, in effect, zero.

The Chair: I do not think that assertions such as that are universal. We

have heard from witnesses who have a different view.

Matthew Sinclair: What I am getting at is that the volition you are talking about—the willingness to license—reflects the fact that, if AI companies try to strike a licensing deal with every web page on the internet, it is obviously impractical. So they are having those licensing conversations around outputs—that is, distinctive data sets that are not publicly available. They are not having them around information that is publicly available and on which they can legitimately train. I would just draw that distinction on where companies are willing to license. They are licensing where they should be doing so under the traditional principles of copyright law.

Antony Walker: I completely understand that the committee's focus is on the UK's copyright regime but the question is: how does that regime function, in relation to AI, in a global market for AI where AI services can be developed anywhere around the world and then deployed into different markets? At the moment, as my fellow panellists have explained, the US has a particularly permissive approach to copyright—for the purposes of innovation, it allows companies to train models on the basis of fair use—and the result is that that is where model training is taking place.

I read a *Press Gazette* article earlier today. It listed 36 significant licensing deals that have taken place in the past year between some of the major model developers and the large rights holders; most of those deals have been done in the United States. At the same time, there are some 20 significant lawsuits taking place in the US; litigation is happening that is testing out the nature of that fair use principle and how it applies in relation to model training and model development.

So we have at the moment a situation where there is a combination of a permissive regime that enables innovation to take place; questions about the extent to which that regime provides a full basis for what is being done, which is being tested in the US courts through litigation; and, at the same time, significant deals being done between significant rights holders and significant model developers. That is the international context of what is happening.

The really interesting thing, which the Government are trying to get at, is: how do we make the UK environment relevant in this global context? How do we create an environment where smaller model developers, technology companies and rights holders can also participate in this marketplace? Our view is that the Government's approach, which they set out in their initial consultation—particularly in option 3 of that consultation—is really about trying to create a market-making mechanism that enables the various parties to get to a point where they can come together, express their rights and negotiate. The three key parts of that package are the TDM exemption, the opt-out—

The Chair: It is not helpful to go over this. We understand that, but the Government have had a reset and moved on to a different position.

Antony Walker: I am here to give evidence on our position, which is that that remains a relevant way to move forward. The third part is the transparency mechanism, which is essential.

Our view is that we need to get those three parts in place to get to a point where rights holders—in particular, small rights holders and individual creators—can express their rights but where there is also potential for mechanisms that enable them to seek licensing deals, perhaps through collective parties and so on. We think that there is a mechanism to get to that endpoint. Our question remains: what is the outcome that we are trying to get to and which is good for the UK, the creative sector and the UK tech sector? How do we get there? That is the fundamental question.

The Chair: I am going to have to ask you to be brief because we have only an hour; we have quite a lot of questions to get through.

Q59 **Viscount Colville of Culross:** I want to ask Matthew to clarify the definition of “learning”. You said that AI companies should not have to pay for training on data because it is learning. I understand what learning is; if I read a textbook, I learn. However, if the models that are training on the data create an LLM, that is something that can be sold on and used. It is expensive to try to do it but, in the end, surely it can be commoditised. Therefore, that “learning”, as you called it, should not be free or without any cost to pay to the content owners for the data.

Matthew Sinclair: Ultimately, I do not think that the relevant question is whether this activity is commercial, in part because non-commercial and commercial work is often a huge grey area. One of the things that policymakers want from the university sector, for example, is that its non-commercial work should support commercial work. So I would not draw the line there.

The line is between the inputs that go into your work. We all learn as we read, engage with the world, have this conversation or watch this session being streamed. When we go out to do our work, whether we are selling that work or providing it for free, we do so on the basis of what we have learned. If what we do is duplicative, there absolutely can be a copyright claim—just as people can breach copyright using traditional editing tools, people can breach copyright using AI tools—but, when it is learning, the AI tool’s attempt to understand the world by ingesting publicly available information would not, if it were a person, trigger a copyright claim. Whether they were going to go and do commercial work or voluntary work cannot be the standard.

Q60 **The Chair:** Can I stop you there? Last week, we had a witness who worked through with us how different models are trained. In some cases, it was a very large number of books; in other cases, it was a very large amount of music. That is not publicly available; in this country, it is available through licensing and through copyright.

Matthew Sinclair: I would draw two distinctions there. One issue is whether or not something is subject to copyright, which is often very

unclear. There is no central database of what is copyrighted and what is not; often, the different copyrights in a given work are complex.

Then there is the question of whether something is publicly available, which means what it says on the tin: can you go and read it, or is it something that is, to put it crudely, not on the internet? I absolutely understand that there have been questions around how people access information that is not publicly available—that is where licensing deals have been taking place—but publicly available information, whether or not it is subject to copyright, is different. I read copyrighted information all day; it is undoubtedly part of why I am able to do the work I do and charge for. I do not think that charging is what determines whether or not something triggers a copyright claim; it is about whether it is a copy versus whether you have learned and then been able to produce original works.

Q61 Viscount Colville of Culross: Are you not making copies to create the database? We have been told by copyright lawyers that you are making a copy of the data, which, therefore, opens you up to copyright claims.

On the idea that it is too complicated to go and find out whether there is copyright, I work in television. It is jolly complicated trying to find the copyright, but you have to do it because the onus is on the channel to make sure that everything is copyright-cleared.

Matthew Sinclair: There are two parts to that. First, on the complexity of it, absolutely. That is why, if you say to companies that have managed to produce tools of phenomenal value by going to billions, potentially trillions, of sources, that they need to do what you had to do when you were in the TV industry, the answer is, “That’s impossible”. Therefore, if we want these tools to be created—or, if we do not, other countries have decided that they should—we need to have a regime that will make that possible. We need to be careful of saying that, because in the context of an individual copyright claim it is already, as you say, phenomenally difficult and trying to do that for a large language model is impossible.

The Chair: We have heard that it is not.

Matthew Sinclair: To your point, there is non-expressive copying in the course of this. That is why a text and data-mining exception is understood to be needed, because in the course of these models they are making copies. That is not expressive and, generally, this is not being shared with the world and therefore is not creating what is properly understood as a copyright claim. That is why Japan and the EU have created text and data-mining exceptions and the US has got there via fair use. I do not want to get stuck on fair use, because fair use is a means of getting to an outcome that every major economy has got to.

The Chair: This is not helpful, because that is not our copyright framework.

Vinous Ali: Can I just interject here to make one point? The way that copyright works on an international basis is that it is a national regime, and Matthew is right to try to segment out these issues. The big elephant

in the room is that this depends on where these models are trained. Currently, they are trained, lawfully, in regimes where this is permitted.

The Chair: I think we understand that.

Vinous Ali: Sorry, I just want to finish this point. It is important that, if we want to create UK AI start-ups, we need to say, "What is happening elsewhere and how can we ensure that we are creating the right conditions to allow these start-ups to thrive and scale?" At the moment, they are not training here. One of the reasons why the case against Getty fell is that Stability AI, which is a UK company, was able to demonstrate that none of its training took place here in the UK. It happened in the US in a lawful way under a regime that is permitted there. As Antony says, there is litigation happening there that has not run its course, but that is what we are up against and we need to decide whether we want to be makers or takers.

The Chair: Can I move on please?

Lord Knight of Weymouth: We are also here to protect our creative sector, which is of massive multi-billion-pound value to our economy. So, we want to achieve both.

Vinous Ali: It is not a zero-sum game, as some would suggest.

Q62 **Lord Knight of Weymouth:** Matt, I want to go back to copyright not used for learning but for output. On material that is used for learning, if I read a book, for example, someone has paid for it and the author and publisher get some money back for that. If I watch television, the licence fee or advertising has paid for that. If I read a newspaper, I have paid for it at the newsagent or online. That is not happening in the learning that the AI companies are doing. They are not paying for it. Therefore, the creatives are losing out. Do you disagree with that?

Matthew Sinclair: Yes, I do. In terms of a TDM exception for lawful access, which is what I see as being under debate, if there is an allegation of unlawful access, that is a question for the courts.

Lord Knight of Weymouth: We saw that with Anthropic and we have had allegations of pirated material.

Matthew Sinclair: But in terms of lawful access, either it is information where there was not a charge or information where they paid that charge. But if you want to say that a TDM exception absolutely should not cover content that has been accessed unlawfully, I do not think anyone here is going to dispute that with you. The payment does not change that. I do not think that what people are looking for is that you pay the paywall charge once to learn. I would separate the two.

Lord Knight of Weymouth: Matthew, all I was trying to establish was that if you are trying to draw a parallel with learning and saying that learning is free, it is not.

Matthew Sinclair: Okay, but learning does not trigger a copyright claim.

Q63 Lord Knight of Weymouth: I know. But there is a remuneration issue and we are concerned about the remuneration for creatives. At the other end of your statement—that output is the point at which copyright comes in and therefore there will be remuneration for copyright—do you believe that there is a technically feasible way of looking at the output of AI models and successfully attributing the source material so that the copyright owners can be remunerated?

Matthew Sinclair: Whether or not the output is a copy should not depend on the source material. The material that will determine whether or not an outputs-based copyright claim is legitimate and successful will be the outputs; the outputs are what you use to identify that. The transparency proposals, which I understand are part of the discussion today, are there to police inputs. That is what people are looking for; it is not that it is necessary in order to prosecute. That is where we have seen lawsuits, when outputs have seemed too close. Sometimes they have failed and sometimes they have succeeded, but they have stood or fallen on the outputs, which are there—you can get them. You do not need transparency for that.

Lord Knight of Weymouth: I do not want to take up any more time. I was trying to establish that the seductive logic of “Copyright is not used for learning, it is used for output and, therefore, by implication, that is where we should focus the monetisation for the remuneration of creatives”, collapses under scrutiny.

Q64 The Chair: We really need to move on. We are getting behind where we should be in our questions. Before I leave this point, and it has been useful, can I ask each panel member to say what the significant factors alongside the copyright framework are? We have talked so far only about the copyright framework. We are talking about investment here and things such as access to compute, the energy crisis, capital and skills. Can you quickly say how significant you think copyright is, alongside those other factors? What do you think is the most important thing? Do we need lots more data centres, or a lot more skills, or a lot more capital?

Matthew Sinclair: I do not think they are independent. One of the reasons why we all feel strongly about copyright is that it would be a huge shame if other work that the Government are doing to address issues such as energy costs and skills were undermined by not having the right copyright regime. We commissioned a study on this point, which found that the TDM was worth about 20% to 40% of AI investment as a stand-alone policy question. Obviously, there are many other factors, but the UK has both advantages and disadvantages in competing for AI investments. You weigh energy against research strength, for example.

The Chair: But 20% to 40% of what?

Matthew Sinclair: Of otherwise anticipated UK AI investments.

The Chair: Colleagues made the point earlier that we have a very

successful creative sector that contributes to the economy and of which we are trying to balance the needs.

Matthew Sinclair: It is important to say—

The Chair: Let me finish. It is an important part of the question—20% to 40% of what? We can look at the creative sector and understand that it brings in a certain amount now. You are talking about something that is possibly in the future.

Matthew Sinclair: It is important to say that the Japanese, EU and US creative industries have all thrived with the protection of the sort we are describing in place. I am not aware of any evidence that suggests that the Japanese text and data-mining exception has compromised the Japanese industry. It is logically hard to see how it would, because the question of whether AI models are trained here in the UK or in the US does not change things for the creative industries. It matters for the UK AI industry, and the best estimate that we have is about 20% to 40%, dependent on this policy change.

The Chair: But 20% to 40% of what?

Matthew Sinclair: Of UK AI investment.

The Chair: Which would be what?

Matthew Sinclair: It is about £1 billion to £2 billion per year in investment, equivalent to about 2% to 4% of total business R&D.

Q65 **The Chair:** Okay. Vinous do you have anything to add, because we are running out of time?

Vinous Ali: Yes, absolutely. This is a critical point. The UK Government have recognised just how important and transformative AI will be. That rests on multiple things, including energy costs, infrastructure, data centre buildout and sovereign capability—we have seen £500 million go into that, £200 million on compute allocation and £100 million on an advance market commitment. The point here is that copyright acts as a ceiling to our ambition. If we really want to become an AI superpower, it is not good enough to have just one of these elements. We may have the best researchers in our universities but, frankly, if they operate on a TDM exemption—which they currently do today on a non-commercial basis—and they want to spin out to create the start-ups of the future and, therefore, want to commercialise this work, they are currently prevented doing so because our TDM exemption does not stretch to commercial. We can have the best research but that will not be enough. We can have the Government pouring £2 billion into AI and unblocking planning laws, infrastructure and energy costs, but that will not be enough if start-ups do not have access to the data they need to create.

I really would press against the point that it is a zero-sum game in terms of choosing between the AI sector and the creative sector.

The Chair: We did not say that it was.

Vinous Ali: The creative sector has put out this number of being worth £124 billion to the UK economy. Some 40% of that—today, not in future—is made up of IT and computing services. Almost half of that creative industry number is represented by the UK's tech sector; one goes hand in hand with the other. Where we have CGI, it has bolstered our film industry; the advent of photography and other innovations have bolstered our creatives and allowed them to achieve more. Distribution and the whole content creator economy have been enabled by tech innovations that have gone before. It is really disingenuous to separate out these two sectors and pit them against each other.

The Chair: We were not doing that. We need to be much quicker with the answers.

Antony Walker: I will be brief. Copyright is, alongside the other issues of compute, energy prices, capital and skills, of pretty equal relevance. My concern is that, if we make the UK a less permissive environment for innovation, copyright will become a bigger issue because it will create even greater incentives for companies to train overseas.

Q66 **Lord McNally:** That is a good intro to what I am trying to think my way through. All of you have given some hard, passionate advice today. If we become obsessed with this concept of the creative industries and putting in all of the various protections, when the great race starts, we will be left on the starting block because of the various protections that have been put in. Getting this right is going to be really important. It is important that, as you say, we do not become overly obsessed with the warm and cuddly creative industries as against the harsh, unfeeling and cynical technical industries. How effective are the existing right reservation mechanisms in balancing the needs of right holders and developers of different sizes?

Vinous Ali: Can I come in on that? First, we should be obsessed with our creative industries. They hold remarkable soft power for the UK. My argument is that AI allows them to go further and faster and allows us to export that soft power. I want to see that be deployed right across the UK economy then globally, rather than having that next iteration of our creative sector happen elsewhere. That is the first point.

The second point is on your specific question about rights reservation. At the moment, we have robots.txt, which is basically a binary system; it is crawl or no crawl. That does a disservice to the creators who are putting their material up on the web in order for it to be searchable and discoverable. There are emerging standards that are coming through to create a more nuanced regime whereby you can say, "Yes, I allow X crawler in for search or for training purposes", et cetera. Those regimes and standards are early on. I know that you will have heard from others who say, "Yes, we have this solution here"—the nature of start-ups is that you always want to project power and the ability to do things—but they do not have the scale that is necessary today. They will get there in time. I would personally like to see that be allowed to run its course, but

it should not hold us back from pursuing a fuller TDM exemption for commercial purposes today.

Matthew Sinclair: The way in which I would delineate is that rights reservation exists as a practical tool today. Reuters Institute found in 2023 that 48% of news publishers were already opting out of the Open AI web crawler, for example. It already exists. The responsible large developers already provide more granularity than is allowed in robots.txt alone; Apple and Google, for example, already allow extensive additional granularity on what you want to allow or not allow.

However, as AI moves forward and is deployed and trained in new ways, we want to be pushing the bar in terms of what is possible. There is an interest on both sides to make that more granular because the AI sector is not going to benefit if people are opting out wholesale. If we can give more granularity, that is in our sector's interest. That work is going on; crucially, it is going on at the companies that are best able to push it forward. What we want to be careful of is going down the road of a solution that is too one-size-fits-all and prevents innovation among smaller AI developers that are never going to be a great source of licensing revenue for the creative industries if you reserve those rights. It is also a huge burden if you are asking them to do complex, additional stuff and to go beyond robots.txt.

The EU's solution is, at a high level, one where compromise can be found: a machine-readable rights reservation should be respected, but do not try to get too specific about how that works in practice.

Q67 **Lord McNally:** I know that other colleagues will want to come in but what role, if any, should the Government or regulators play in getting this?

Antony Walker: First, in terms of rights reservation mechanisms at the moment, we have seen a significant increase in their use and deployment; to me, this suggests that there is some value in them. However, we recognise, as colleagues have said, that they are still a blunt tool that requires considerable refinement. That refinement is happening through international standards bodies, which are the right entities to do this complex technical work because, in essence, this is about the fundamental standards and protocols that underpin our online world.

Where the Government can help is in signalling. That is why we think that the rights reservation approach expressed in option 3 of the consultation was a positive signal: it pushes more emphasis on getting that work right. As I said, getting that model of TDM rights reservation and transparency requires you to get these mechanical tools right. There is definite progress being made towards a much more refined set of tools.

Again, this is about market-making. How do we ensure that the tech sector thrives in the UK; that the UK economy thrives in a world of AI; and that the creative industries thrive? We have to create a marketplace

where they can come together and trade value. There is absolutely huge, enormous value in the creative industries. Our perspective is that the proposals put forward by the Government were dismissed too quickly. There was not enough proper consideration given to how they could provide a workable mechanism to get to the outcome that we all want: a thriving UK economy, a thriving UK tech sector, thriving UK creative industries and a society that has creativity in the arts at its core.

The Chair: We would all like to see that.

Q68 **Baroness Wheatcroft:** Can you tell us—in the world that we are envisaging, where there is some protection for the creators of unique works—what you thought might work in offering them some sort of protection, even if it was across industry groupings?

Antony Walker: Again, the fundamental point is creating a market, because there is fundamental value there. If you help create an easy way for rights holders and creators to express their rights, that then drives people to a market where more licensing will happen. Crucially, more licensing will happen in the UK because you have the TDM exception, which encourages model training to happen in the UK. So if you want this to happen in the UK, you have to get the model training working here.

I also stress that we should talk about small models as well as large models. Often, it is suggested that energy costs are prohibitive. That is less the case when you are looking at small models or refinements to models, which can be of equally high value; a small bit of content can be incredibly valuable for refining a model and so on. So the question is: how do we create that kind of marketplace?

We have always subsidised the arts in this country because we see their wider public good aspects, but that is a separate question from the copyright question. It is important that we—I know it is really difficult to do so—slightly separate out the question of how we ensure ongoing funding for the arts and creativity from issues around how people express their rights and get remuneration for their rights. There is a slight separation when we are talking about copyright as it applies to AI model development and model training. But I reiterate that there are lots of technology companies that sit right on the cusp of these issues: they are rights holders and model developers, and they also want to find a route through to this outcome.

Baroness Wheatcroft: That is certainly the case. Vinous, would you like to add to that at all?

Vinous Ali: I will just build on what Antony said. If there is a broader TDM exemption, I think that the pie will essentially grow bigger for everyone. A TDM exemption would allow for further innovation and, importantly, draw back some of that training and innovation to ensure that it happens here in the UK and that value accrues to the UK economy and society.

Q69 **Baroness Wheatcroft:** So, when you say that you would like a broader TDM exemption, do you want to extend it, as you said earlier?

Vinous Ali: I want to extend it to commercial purposes—absolutely. Then the question turns to licensing markets et cetera. From the start-up perspective, the question is whether those start-ups can compete with those who have larger, deeper pockets. Therefore, what I would warn against is policy and government legislative practices running ahead of where the technology is today. If it is not something that is accessible to UK start-ups, if it does not deliver on the granularity needed for individual rights holders to be compensated rather than large publishing companies—Paramount is currently putting, I think, \$108 billion on the table for Warner in cash; these guys are already well covered—and if it just becomes a game between large model holders and large rights holders, then the UK loses. For me, it is about making sure that the Government do not legislate ahead of where the technology is today to deliver for UK start-ups and UK creatives at the individual level.

Matthew Sinclair: I think that that is where principle is helpful. The closer you get to the specifics, the more your system is going to unravel as the technology changes. Whereas, if the principle is a text and data-mining exception to ensure that copyright claims are focused where they should be—and then a rights reservation, which should be not only machine readable but align with industry standards over time rather than being too specific—you can have protections that will move with the market, rather than, frankly, us all coming back and doing this again in a few years. I suspect that that is the risk if we get too specific.

Q70 **Baroness Healy of Primrose Hill:** I will return to transparency because that is one of the fundamentals. What specific information can developers provide to support meaningful transparency for rights holders? Promoting greater trust and transparency between the AI sector, all levels of it, and the creative industries is one of the key objectives of the Government's AI and copyright consultation. Matthew, you mentioned the EU AI Act, so maybe we could start with you.

Matthew Sinclair: My colleagues have done a lot of work on this in Brussels. First, it is important to understand what transparency in the EU constitutes: companies are required to provide a summary. There is a debate over what that summary should include but, crucially, it is not source-level information. I will not go on about the problems of source-level transparency; that is the point at which this becomes impossible. But the nature of that summary is reasonable; I am happy to explain why, but I do not want to impose on your time.

When we are talking about a summary, there is a reasonable debate to be had. CCIA—working with our members, which includes many of the largest AI developers—published a set of principles for AI summaries to contribute to that in Brussels. I will not list the principles here, but they are trying to ensure that we do not undermine AI security, which overly granular transparency requests will do—they will provide an instruction manual for poisoning these models—and that we do not undermine the

dynamism of the sector. One of the risks is that, if you insist that everyone has to show their homework in terms of the data they have used, you will lose the ability for bright start-ups to come along and outcompete people on the basis of using data in new and innovative ways. You will lose what is currently this amazing dynamism in the sector. So long as we are doing something that is technologically feasible and not pushing into that confidential and security-sensitive business information, I think there is a reasonable discussion to be had around the summary information.

We also have to step back and look at what it is for. If we have a TDM exception requiring extremely onerous transparency requests to show a load of work that is legal, that will not achieve very much for rights holders. If we do not have a TDM exception, no one is training here anyway, so no one is filling out our transparency forms. So we need to be clear on what this transparency is for and then come to a reasonable summary that people can deliver without compromising their businesses and their ability to develop models here in the UK.

Q71 **Baroness Healy of Primrose Hill:** Ms Ali, what do you think about this for scaling up? Can the smaller companies accept the principle of transparency and still work with the creatives?

Vinous Ali: The question is: transparency for what and whom? There are many reasons why transparency is important. For example, in the UK we have the AI Security Institute, through which a lot of testing of pre-model deployment happens between the model developers and the experts within AISI for reasons of security and ethics.

The Chair: This committee is concerned about the remuneration of creatives. We are very short on time, so let us not get into other issues. This is not about security issues; we are concerned with the point that we made at the start.

Vinous Ali: We are talking about the transparency to enable remuneration at an individual level. That requires a granularity that is impossible for start-ups to meet. Some of these models are trained on billions of tokens and pieces of data. There is no copyright register that says, "Here are all the copyrighted works. Please check against them". Even holding that data would cost a great deal of money just in terms of hosting, and then there are the costs to make it searchable.

There is another question. I think that the EU has gone for 10% of the top domains. That is not to say that those domains have been lawfully accessed but that there is a copyright claim somewhere else. I think that one of your previous witnesses touched on this point. So if you are not able to offer transparency in order to meet the individual, granular remuneration for rights holders, what is the point of it? Does that just lock out UK AI start-ups that do not have the ability to pay those costs?

Q72 **Baroness Healy of Primrose Hill:** Mr Walker, you talked about the market-making mechanism. Where would transparency fall within that?

Antony Walker: I think it is part of the mechanism; I do not think it is the whole solution. I have heard it referenced as, "If you just had transparency, then everything's solved". However, I do not think that is necessarily the case, because there are real challenges to providing a level of granularity to the transparency that is useful, particularly for smaller rights holders.

If I look across our membership—we have a very broad membership—transparency is the area where there is probably still the widest range of views at the moment. I take from that that it is still an open area for development and discussion. You have to think about what you are trying to deliver with this transparency mechanism. Are you trying to deliver a list that every author or songwriter can go and check? Is it at that level of granularity or are these the broad datasets on which these models have been trained? There are questions about that.

The other thing you have to recognise is that you need to think through issues of trade secrets and what happens if you give away a whole load of proprietary information. That is a fundamental disincentive for model training in that jurisdiction. If I have to give away all the fundamental IP about how my model is built if I train and license in the UK, but I do not have to do that in the US, what would my legal counsel say? My legal counsel would say, "Go and train the model in the US". So I think it is part of the solution to this challenge, but not the holy grail. We can make progress on it to find something that will be useful, but it is not the endpoint.

- Q73 **Lord Knight of Weymouth:** Can I test a thought with you? A lot of this conversation is rooted in thinking that is a bit old—from when the development of generative AI was emerging. We are in a more mature phase: we are seeing a lot of slop and a lot of quality problems with AI, so there is much more of a market now for training using higher-quality data. That is why in a country like Japan, with a big pushback from the creative sector, you are seeing big licensing deals starting to emerge there, because they want quality in their AI output. So, to an extent, is this whole TDM thing, and some of the arguments that are being made, a bit of a distraction? In the end, as we move towards higher-quality AI products, we will want higher-quality data, so licensing deals will therefore be done.

Antony Walker: That is an excellent question. The TDM exemption remains relevant for a couple of reasons. First, there is lots of content out there that is really valuable and does not need to be copyrighted. Creating easier access to that data through an exemption just makes the process of model training much easier and more reliable and predictable from a legal perspective. So the TDM exemption is helpful for lots of data that is really of no value to anybody other than model developers.

- Q74 **The Chair:** Could you give us an example of that?

Antony Walker: For example, really blurry images and photographs can be useful for training visual models. They are used in autonomous

vehicles; being able to interpret those blurry images to a huge scale can be really valuable when you are training an AI to understand where it is and so on. So that data would essentially be useless to anybody else, but is really valuable in that context. However, trying to license that would be really difficult.

Then again, once you have the TDM exemption combined with a rights reservation process that provides that opt-out, it would more clearly create a market for those deals to be done. My argument would be that, yes, you are seeing more deals being done in Japan, but I think you will see even more deals done in the UK if we have that combination of the exemption in place, the opt-out and some transparency mechanisms that are workable and agreeable between the various parties. I would like to see us explore those issues more. Over the last year, we have not been doing that; we have been in a standoff between the two parties.

The Chair: I come to Viscount Colville, because he has to go soon.

- Q75 **Viscount Colville of Culross:** Sorry; I want to keep following up on the transparency thing. Matthew, you said that the EU AI Act offered a possible way forward with the transparency, by having a higher-level summary of databases. However, I read that Meta has said that even the EU position is overreach and Google is really worried about it being too demanding. We are certainly being told by AI developers that the regime in the EU has not made it very attractive to set up there.

Matthew Sinclair: Sorry; to be clear, the EU's approach of a summary could get you to a good answer, but the EU's implementation of that is really struggling. UK policymakers should be very careful about thinking that the EU is further along than it is, as it has not managed to land this yet.

How should I put this? It is also good to be clear that the EU is not trying to get to source-level transparency of the sort that has been discussed here. I think that the EU is going too far within the realms of a summary solution, which is more tenable. If the UK wants to go down this route of transparency—again, it is not 100% clear what that is achieving for people—a summary is the only way that it would be practical. If we go down the route of source-level information, we would literally be talking about the largest regulatory disclosure in history to create this system—almost by definition, given the scale of the data with which these models are working. That is clearly untenable. So, within the broad space where the EU is looking, a summary—along with a TDM exception, which is obviously the starting point in the EU—is more tenable. Within that, there absolutely are problems with the way that the EU is doing it.

- Q76 **Viscount Colville of Culross:** I know that we are in a hurry, but we absolutely need an answer to this, because so many of the content creators who have come to us have said that, at a granular level, a URL is necessary to make sure that we get proper remuneration and weighting for the way that their content is used. We were told last week that developers necessarily keep detailed records of databases used to

build their models, and that it is technically feasible to require them to retain and disclose information to reproduce those training databases. Is that not true?

Antony Walker: Can I challenge that? Look at the US: there are no transparency requirements in the US, but I listed a significant number of very large deals that have been done with very large rights holders and LLMs. If it is impossible to come to a valuation because you do not have that level of granular detail, how could they have come to the point at which they made those commercial deals? How could their boards have accepted them as sensible things to do? I question whether you really need that level of granular transparency when, in other markets that do not have that level of transparency, real, big, commercial deals have been done that are based on valuations. My suggestion is that there are other ways to come to those kinds of market valuations.

There is a question about whether smaller rights holders have the ability to do that. That is where the whole question of collective licensing comes in. One of the real goals of what we are trying to achieve in the UK should be focused on how to create a market mechanism that works for small rights holders, individual creators and small technology companies.

Matthew Sinclair: To answer the question, I think that that statement was wrong. I think it was based on an overly narrow view from someone who had been involved in the development of some models, but who has not fully thought through what it would mean to turn this into a regulatory requirement. It would require companies to produce billions if not trillions of URLs for models that, essentially, are trained on the internet. It is not clear whether it is possible to meet the technical feasibility challenges, let alone at a cost that makes it possible for the broad swathe of AI developers.

It would also create enormous security risks. There have been important studies on data poisoning recently, and you would be making almost all solutions to that very hard to achieve.

Q77 **Viscount Colville of Culross:** You have made that point already. Thank you very much—I am sorry to interrupt you, but we need to speed along. Antony raised the market mechanism for licensing, which would allow the large content providers, as well as the smaller content providers, to get some value from their data being used. We are seeing these licence deals and have talked about them extensively; however, we keep hearing from small and medium-sized content providers that they do not work for them. Is there a role for government to help set up an effective market mechanism, so that all content providers can be incorporated and licensed, or should we just leave it to the market? Vinous, could you help us?

Vinous Ali: It is a really difficult question. We have spoken to start-ups that have tried to strike licensing deals with some of the smaller rights holders, understanding that they may never get access to those larger rights holders and publishers. They do not get a response because they are holding out for deals with larger language models. This question

around transparency and the granularity of that is where the UK might win in the AI race at the application layer. That is: are we going to build the next OpenAI? Perhaps there is an outside chance: costs are coming down and there is definitely the talent here in the UK, but it is an outside chance. The application layer is where things get interesting. That is essentially building on those foundation models.

There is then a question around liability. That is, if you are building on these models here in the UK, you may be held liable to information within those models, but you cannot test it because, frankly, those models were trained in the US where your transparency regime is never going to reach.

It then becomes a question of liability. One of the things that we sometimes miss in this broader debate is that, if we want to see AI diffused right across the economy, which I hope we all agree we want, that question of liability becomes even more important. It matters because the UK has lagged in other technologies in terms of broader adoption. We in the UK sit somewhere around the 24th or 25th in terms of robotics density; we are slow to adopt cloud, et cetera.

I do not want us to lag behind but, if we create uncertainty not only for the developers of AI but for those who are deploying and building on top of these foundation models, essentially it could crumble. Then if you open up the Pandora's box of, "These foundation models will have to comply with whatever transparency obligations that we in the UK put on them for reasons of trade secrets or data poisoning", and so on, and if we are saying we do not want to see these models deployed in the UK, then we will frankly be behind. We are through the looking glass there.

Q78 Viscount Colville of Culross: Antony, you raised the idea of market mechanisms for licensing. Do the Government have a role in trying to set up those licensing exchanges or will the market just do it and make sure that everybody is included?

Antony Walker: I was speaking last week to some representatives of organisations that represent individual creators and small rights holders. One of the points that they were making was really valid. It was that they need to build confidence among their constituency that you can license. One of the results of the debate that we have had through the course of this year is that some of those individuals almost thought, "That is it. I'm done. It's finished. There is no way for me to have any agency in this kind of environment at all". So there is something about how we bring those people back into the conversation and create collective licensing agencies that can bring those interests together with the market.

We also have to recognise that we are in the middle of a very dynamic moment where prices are not set. It is difficult to know whether you have content that is incredibly niche and therefore irrelevant, or content that is incredibly niche and incredibly valuable. A lot of people do not quite know what that situation is. At the moment, we are in this shaking-out point in the market, where we are starting to understand what the

relevant relative value might be of different types of content. Unfortunately, there might be some types of content that are not particularly valuable in this kind of AI model-training environment. But there will be other things that, as I said, are niche that could prove to be extraordinarily valuable. We need to unlock that in the UK.

Matthew Sinclair: We should not assume that these different kinds of models are substitutes. Often, they are complements. The model trained on a big old pile of dumb data often works with the model that works on specific niche data. Your question, Lord Knight, is relevant on this point. They are not separate things. These different models are working together to produce an output. The big LLMs' role is not going away; it will be working with other applications built on top. It is not going to expire.

Lord Knight of Weymouth: There could be a regime that is about how we go on from here, and then we will worry about the past differently.

Antony Walker: That is right. My perspective is that the US courts will determine the lawfulness of what has happened in the past and any liabilities. That process is playing out with at least 20 significant lawsuits going through. The cumulative effect of those cases and the judgments in them will define the question of what is learning. That is where that issue will fundamentally be defined.

What we are trying to talk about, and what the Government were trying to talk about in their consultation was: what do we put in place going forward and how do we unlock the potential? The UK, in a recent global index, was rated fourth in AI innovation. That is a phenomenal position for the UK to be in. It gives us the opportunity to shape a world that will be dramatically reconstituted by AI. We have the power to do that. If we can get our approach to this issue right, it will be one of the ways in which we can shape what the AI future will look like internationally. But to do that, we have to get everybody around the table. Unfortunately, over the past year, there has been a megaphone approach to policymaking rather than a constructive one.

The Chair: We need to stop there because we have run out of time. Thank you very much for your input today.