HOUSE OF LORDS

Communications and Digital Committee

Corrected oral evidence: AI and copyright

Tuesday 13 January 2026

2.05 pm

Watch the meeting

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

Evidence Session No. 6 Heard in Public Questions 106 – 134

Witnesses

I: Roxanne Carter, Global IP Lead, Government Affairs and Public Policy, Google; Guy Gadney, Chief Executive Officer, Charismatic.ai

USE OF THE TRANSCRIPT

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

Examination of witnesses

Roxanne Carter and Guy Gadney.

Q106 **The Chair:** Good afternoon and welcome to this meeting of the Communications and Digital Committee, which is our first meeting of 2026. My name is Baroness Barbara Keeley, and I am the chair of the committee. We will be taking evidence from two sets of witnesses this afternoon in the final sessions of our inquiry into AI and copyright. I am pleased to welcome our first panel, who are representing Google and Charismatic.ai, and I particularly want to thank them as other platforms we asked to attend today did not feel able to do so. I am pleased that you two are able to come and do that. Today's sessions are being broadcast live and a transcript will be taken. Our witnesses will have the opportunity to make corrections to that transcript where necessary. Can I ask if you would introduce yourselves to the committee, starting with you, Roxanne Carter?

Roxanne Carter: Thank you for the invitation. I work at Google in its government affairs and public policy team, and I lead on AI, copyright, news and media policy.

Guy Gadney: I run an AI studio based in the UK called Charismatic.ai, which has a generative story engine that we have developed over the past couple of years alongside partners Channel 4 and Aardman Animations, and is funded by Innovate UK. I also sit as a trustee on the Story Museum in Oxford, looking at literacy among kids, and on Sheffield DocFest.

Q107 **The Chair:** Our first question to open the discussion is key and central to all our deliberations on this committee. Would you agree that UK creative industries' rights holders should receive payment or revenue share when their works are used by AI systems? I start with you, Roxanne.

Roxanne Carter: This is an important and timely inquiry. From where we sit as Google and YouTube, we are an AI model developer but we also have a long-term history of partnership with the UK creative industries. We believe that the AI opportunity for the UK is significant but we need an enabling copyright framework. This is not just impacting Google and Google DeepMind, which was founded in the UK with cutting-edge research and development; it will have an impact on all other markets and sectors across the economy where we want to be developing our own AI sovereign capabilities. We also want to ensure that UK users have AI tools that speak to British identity and culture, and we want to make sure that the UK has a role on the global stage when we are thinking about AI development.

We need Parliament to strike a balance to ensure that the UK creative industries can continue to flourish while enabling AI. It is in that spirit of balance that we look at this issue of payment and remuneration. When it comes to training AI models on content that is freely available on the open web, we do not believe that we should license that content. What

the AI model is trying to do is analyse huge amounts of data to identify patterns and statistical relationships between words and language concepts. It is not an information retrieval system. It is not a database. It is not looking to make copies. It is trying to develop new tools to then produce wholly new content.

However, we are seeing a market develop for the access to content. What do I mean by that? It is for archived content, for specialised datasets, for content that may be off-platform in some way or for opted-out content. We perhaps might come to that later on. From where we sit, we think that having a text and data-mining exception here in the UK will help facilitate more licensing arrangements for access to content.

Q108 **The Chair:** You are not really answering the question that I put to you. Do you think that creative industries rights holders should receive payment or revenue share? It is important that you answer the question.

Roxanne Carter: Let me be clear here. It depends on whether you are asking us to pay for every single piece of content for the training of the model. If that content is freely available on the web, then no. But if you are asking us whether there should be deals for access to content that might be off-platform or archived, then yes, absolutely. Those deals are being done.

Guy Gadney: We have been operating as a small-to-medium enterprise in the UK in the field of AI in the creative industries for over 15 years now, which puts us in a slightly strange position of having seen a few cycles of this conversation going around. To answer the question, from a company that is entrepreneurial in its nature, based in the UK, I look at it through two lenses. One is morally. If anyone uses someone else's copyright material for commercial purposes, yes, there should be financial recognition. That is my personal view. Legally, there is less clarity and we need to look at this through the legal lens of the current status of the precedents. I go back to the US case of the Authors Guild v Google between 2005 and 2015, which used the word "transformative" away from the original copyright material for a new use specifically for training—a case that Google won. From a legal standpoint, I have yet to see clarity that enables us either to operate or have conversations with investors or other commercial entities on why AI systems should pay for copyrighted material in their training functions. There is need for an urgent intervention for clarity from Parliament and the legal system, noting that a lot of the advances in the last week or so have been through legal cases brought, as distinct from legislation.

Q109 **Baroness Wheatcroft:** I should particularly like to ask Roxanne something, on the basis of your previous answer. What often concerns people is that what is trawled for training purposes from open sources with, as you would have it, no payment, can then be used not merely to train but to replicate the work of authors. They fear that effectively their creativity is going to be taken from them without any payment. Before we move on, I should like to know whether there is an issue there and, if so, how it should be dealt with.

Roxanne Carter: Certainly, from the Google side, when we are designing these tools, the purpose is to then create wholly new content. To get to that point, you need to feed the model, for want of a better word, really huge datasets—really massive ones. Then that guards against bias. It ensures different perspectives. I was trying to think of an analogy here. If, for example, in all the material you give the model, you just say, "peanut butter, peanut butter, peanut butter", it will analyse that across the dataset and learn that "peanut" is followed by "butter". So, it does not capture "peanut allergy", and we need to give it as much data as you possibly can.

On that point about replication, the more diverse information you can get, you then restrict the issue of replication, as these models should not be replicating in the first place. There was a very rare case of memorisation in what we were seeing at the beginning when these models were coming out. You have things like output filters and more technical controls that have been able to manage that. What we are seeing now is that that is not happening as much as it originally was.

Baroness Wheatcroft: But the possibility remains, though.

Roxanne Carter: I do not think it does, actually. At the beginning, there was that risk. The models are now performing much better because they are designed to produce wholly new work.

Guy Gadney: As Roxanne says, the market is moving fast. Certainly, if you roll back to a year ago, you had that problem significantly. Remember that there are two types of rights that you are looking at. Again, I am not legally trained, so I put that as a strong caveat. The first is copyright used for training. The second is: does the result look like something that is already copyrighted? Certainly, we have seen a lot of instances of those since generative AI first came online. There are two different categories there. Have we seen instances of outputs, generations that anyone has generated, that would by anyone's standards infringe copyright? Yes, 100%. But as Roxanne said, it takes a while for that technology to evolve to determine whether or not something is copied and therefore has generated something that infringes copyright, and build the technologies to prevent that happening. This is a moveable feast at the moment. Can filters be put in place to stop it? Yes.

Q110 **Baroness Wheatcroft:** How effective is rights reservation? If I really want to make sure that my rights are not infringed, does that work at the moment?

Guy Gadney: There are two sides to this: a rights reservation around preventing collection, and prevention of use. The recently resolved case brought against Anthropic, and a class-action suit brought in the US by a law firm, had two principles, as you probably know. One was around collection and the other was around piracy. Where the class action succeeded was in piracy, not generation. There are elements in this that we are already seeing coming out in various responses.

This goes back many years. The creative industries, music labels, publishers and video companies have made occasional attempts going back to when I started in this industry, in the mid-90s, as the first head of digital for Penguin Books. Various rights reservation tools, strategies and technologies have been put in place but, broadly, these have failed consistently over the past decades. As companies such as Spotify and Apple Music have shown, the solution to this is to be innovative and create a product which fills the vacuum and allows new revenue streams to go back to the creators and meet what the market wants. For reference, by the way, we did research around 2011 into online pirates, which was a difficult piece of research to do, which showed that the majority of these individuals were happy to pay for a subscription of around $10 to $15 per month if they could gain access to the music they wanted. That was prior to iTunes launching.

There is a real challenge technologically, practically and from a will perspective for various organisations to put in place the technologies that allow this, but the solution to it is also very market-driven—what audiences want at a particular moment in time. I point as a positive example of this to the recent agreement between Disney and OpenAI, which will allow more playfulness, in broad terms, with the copyright that Disney owns.

Q111 **Baroness Wheatcroft:** Building on that, Roxanne, is there more that companies could do to make sure that opt-outs are effective?

Roxanne Carter: We take a slightly different view. We believe that the current tools are effective. They are built on robots.txt, which has been a well-known standard for 30-odd years. It is simple, scalable and universally understood. The simplicity here is a feature, not a bug. Because of that, you are seeing the large media sites, for example, being able to opt out as well as a small personal blogger. In having that choice and control to opt out your content from AI training, we see the robots.txt mechanism as effective. The feedback that we have received from the ecosystem that we have been consulting has been positive as well.

Q112 **Baroness Wheatcroft:** There seems to be a fear that, if companies opt out, they will not appear as they should on Google Search. Can you put their minds at rest on that?

Roxanne Carter: Yes, and thank you for the question, because I have watched previous sessions and I know that this has been a concern for the committee. The short answer is that you can remain on Search and yet still opt out of your content being used for AI training. When we were designing our generative AI tools, we launched a consultation and we spoke very widely with the web ecosystem. One of the clear pieces of feedback we got was that they want to remain on Search, but they want to have the control to opt out of their content use for AI training. So, in 2023, we launched something called Google-Extended, which sits in robots.txt but is a stand-alone control that allows you to say yes to Search and no to AI training.

Baroness Wheatcroft: That was in 2023, so why are people still concerned?

Q113 **The Chair:** I just intervene here for those who have not referred back to our earlier session, where this issue was put to us in very strong terms. In fact, the witness told the committee that Google uses its search dominance to have an unfair advantage because it "combines its … search crawler with the crawler that accesses AI and performs AI summaries of contents and live AI searches". They said that, unless this problem is addressed, stakeholders cannot create a fair marketplace for everyone. It is a question of your place in the marketplace, which clearly the Competition and Markets Authority is looking at.

Others said similar things, including that the hold-up in trying to develop a fair marketplace was the lack of participation by Google, because you have that unfair advantage. The comment was made that, if Google is not paying, no one else will either. We also heard from another witness who used to be a Google employee that "Google is a bad actor in this … because if Google does not have to pay for content, then no one else is going to". How do you answer that? Surely those cannot be labels that Google is proud of having.

Roxanne Carter: No, absolutely not. I will take those point by point. Through Google-Extended, you can opt out of AI training and remain on Search. You can see this if you go to the robots.txt controls, for example, for the BBC, the *Daily Telegraph* and the *FT*. What I think the panellist in the last session was getting to was how gen AI features are being rolled out in Search. That is a slightly different issue that is being tackled by the Competition and Markets Authority. That is about how, as the Search project evolves, those features are then being displayed.

Q114 **The Chair:** Just to clarify, they may be able to opt out of AI training, but are you saying that they can opt out of AI overviews, as that affects the number of clicks down the line?

Roxanne Carter: This is a live ongoing discussion and process with the CMA. I cannot give more details at this time, but I would be very happy to write back to the committee. We appreciate that a consultation is due shortly, so we would like to keep you informed on that development.

Q115 **Lord Knight of Weymouth:** We would be really interested to hear about that, because my understanding is that, if you were to remove yourself from the AI training, you have to turn nosnippet on and, if you do that, you reduce click-through by 50%. My concern about leaning heavily on robots.txt is that it is not mandatory and it is at a site level. If you are an individual creator with an individual item that you want to protect, you have to negotiate all the different places where it might appear and what the robots.txt arrangements might be. That feels inadequate. If an opt-out is to work, do we not need one that will work at an item level, not a site level? Would you agree?

Roxanne Carter: One of the reasons why Google-Extended works so

well is that it is not a copyright control, it is a web publisher control. One of the challenges we see is that nothing on the internet is labelled as copyright. There is the ambiguity issue. So it is very hard for a tech platform to be able to identify and then police what is copyright protected, what is an orphan work and what is in the public domain. That makes it extraordinarily complicated. The other issue with asset-level opt-outs is that it will not be clear to the AI model developer what is being licensed to who, under what circumstances and for what time period. This is a huge amount of data that you would have to have in a reliable registry, for example, which simply does not exist.

Lord Knight of Weymouth: Crypto relies on reliable registries and happens at scale. This is not impossible technically, is it?

Roxanne Carter: It could be. I would slightly push back on that, only because, looking at my experience from the broadcasting side as well as the tech, licensing is incredibly fluid. You may choose to license your content to one entity tomorrow and then have another licensing arrangement for the next day. What is amazing about copyright is that it allows you to do that—it is the fluidity. Asking effectively the rights holder to manage a registry and keep it up to date all the time is not just an administrative hurdle; it is really consequential and very difficult.

Q116 **Viscount Colville of Culross:** We have heard a lot of evidence about transparency and the need for it, both at a high level and at a granular level. We know, from your written evidence and from lots of other people who are doing AI developing, how expensive and difficult it is to give granular detail. Do you use third-party data collected for your LLMs? Please give me a short answer.

Roxanne Carter: A short answer is that, in trying to train the models and have as much data as we possibly can, we are doing deals for licensed content, yes.

Q117 **Viscount Colville of Culross:** As that is the case, you obviously have to be careful about the sort of data that you are using. You produced a paper in 2020 entitled "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". In it, you noted specific criteria that were needed for your AI models to clean up third-party data that had been collected. You have very specific criteria for cleaning up that data. The paper says: "Many pages had boilerplate policy notices, so we removed any lines containing the strings 'terms of use', 'privacy policy', 'cookie policy'". If you can get that level of data while training your models, why would you oppose the reporting of granular data as too difficult? Forget about the competition and the security issue; I would like your response just on the sheer difficulty issue.

Roxanne Carter: For us, transparency is important, and it has been a core value for Google. We were one of the first internet companies to publish a transparency report, back in 2010. Specifically on copyright, we regularly publish copyright enforcement for Google on both Search and YouTube. We are trying to wrestle with this in a new era of AI. I worry

that, if you are asking us to provide, URL by URL, a granular list—I think that is where this goes to—you are asking us to summarise the internet. The datasets are just so massive. I think you have seen this in court cases as well: it is trying to identify a needle in a haystack. It is not just that it is administratively difficult; it is that the scale of the operation is just too vast, even for Google.

Q118 **Viscount Colville of Culross:** I quite see that it is an enormous operation to try to get that detail. But if you have this third-party data collected, before you deploy it in your end-user models, such as Gemini, surely you need to be absolutely clear about whether any of this stuff is being taken from the dark web and whether there is any CSAM in it. Surely you need to have looked at a really granular level to make sure that you are cleansing that data properly.

Roxanne Carter: We do. To your point, we look at it in terms of datasets—data categorisation—and that is the level of granularity that we can get to, but that is not URL by URL.

Q119 **Viscount Colville of Culross:** What is it?

Roxanne Carter: It is the datasets. It is the bucket: is this licensed content, crawled content or a publicly available dataset in some way? It is in those broad categories, rather than specific URLs.

Q120 **Viscount Colville of Culross:** Those broad categories will not tell you whether you have CSAM and will deploy it in your end-user models, will they? You need to be clear that that will not happen, do you not? You need that level of detail, do you not?

Roxanne Carter: I would contend that it does, but I am very happy to provide some more written feedback on this.

Q121 **Viscount Colville of Culross:** Thank you very much, because obviously it would be really worrying if you did not have that sort of detail. It seems to me that that is much more than just a broad-brush database.

The other point against granularity that you and many other developers have talked about is that it will create both a security and a competition risk. Various people have come up with ideas for how you can deal with that, but one thing that seems to be most effective is the idea of a black box, which could be held by the regulator or the AI Security Institute. Apparently, this is done in competition law already. You would put your LLM's entire database into a black box, and then the rights holder would be able to get their discrete information that would have come from that LLM. You would not, in that way, be disclosing the whole set, and therefore you would not have a competition risk or security risk. But you would then allow the rights holder to find out whether they have any content that has been included and not paid for in the LLM training. What is your response to that suggestion?

Roxanne Carter: I have not heard this particular proposal, which I would like to go back to the company with. I know that this is a very live

conversation in many jurisdictions, as we try to grapple with what transparency looks like. I know, for example, that we have had very detailed conversations in Europe, where we focused on "a sufficiently detailed summary"—I think those were the words—which is kind of contradictory. Finding new solutions is important and worthy of discussion, but I stress that we are seeing those transparency discussions happen where a text and data-mining exception already exists. That bigger picture of the enabling legal framework to allow for the training is important, and then you look at the transparency solutions that might be available.

Q122 **Viscount Colville of Culross:** You are a senior executive at Google and you have not heard this black box suggestion—I am quite surprised. But, anyway, what is your immediate instinct? Why would that be a threat to either your security or your competitiveness?

Roxanne Carter: I would really need to look at the details. I am happy to do so. If you can provide that to me, I would be happy to take that back and then get back to you with more details in writing.

Q123 **Viscount Colville of Culross:** What would happen if our recommendation was that there should be granular detail of the data collected in LLMs, and regulation was then put out to support that? What effect would that have on Google rolling out its models in this country?

Roxanne Carter: We would have to see the specific details of that. These are live conversations in other jurisdictions as well, and we take this very seriously. We want to engage constructively in debates on transparency. That was part of the reason that we signed the code of practice in Europe, so that we could be part of that collaborative discussion. We would have to see the details and calibrate on what the requirement was.

Q124 **Viscount Colville of Culross:** Guy, I am sorry; I have been rather Google-centric in my questions. This is something that you might not have thought about but has been put to me. There are open-source LLMs—in particular, one is being run by the Allen Institute in America, set up by Paul Allen from Microsoft. They have come up with an LLM called Olmo 3 AI, and apparently it was built for just $90 million. Part of its selling point is that it is open about how it is trained, what the training decisions are and what the data used are. Is that not something we should roll out in this country, encouraging a similar sort of UK open-source AI model based on copyright-cleared UK material?

Guy Gadney: We were talking about transparency—I have yet to see a good definition of that word, especially when it is applied to technology, which can go down into multiple granular levels and is linguistically complex. We need to be very clear what we are talking about. Your last question to Roxanne around whether there would be an advantage in the US versus the UK in terms of access to those models is very pertinent, because we see that it is very important that we as a country can

operate competitively and have access to the latest technologies to be able to compete with our international competitors.

At the moment, we are struggling because overseas territories most commonly have access to Google models and other models before we do as a nation. That may be a short amount of time—only two or three weeks—but in this world two weeks is a long time to be able to get a competitive foothold. I would caution against any form of legislation that would prejudice either the entrepreneurial and nascent creative AI sector and/or the creative industries.

To your point about open and ethically tick-boxed—whatever we want to call it—models, of course we would welcome that. We have been lobbying for a while for a sovereign large language model in the UK that is not only able to have all the data processing done in this country but is trained on data that is sourced and remunerated properly, and on UK copyright and UK material as well. I do not think that having solely UK content would generate an internationally competitive model, but certainly your figure for Olmo 3 of $90 million is indicative of how the comments made previously by larger tech companies that it is impossible to make a new large language model because it is simply prohibitively expensive were false. We are seeing the ability to have more innovation in this sector coming up very rapidly. I would urge this country to do more in that sector as well. We have yet to see something like Olmo 3 in the UK, and there is absolutely no reason why that should not have happened already.

Q125 **Viscount Colville of Culross:** Should the Government be doing something to encourage those open-source models to take place here?

Guy Gadney: Yes, I absolutely believe so. There is an urgent need for an intervention at a significant level, in both the technology that sits on it and the use cases that enable that large language model. As Google and others have shown, it is not just about having a piece of technology; it is what you can produce with that piece of technology that can inspire other people to create with it.

I go back to Google Books v Authors Guild. It was not a commonly referenced case, in my personal view largely because you could not see what was going on. But over the last couple of years, as we have started to see image generation and video generation coming out, people have said: "This looks like Mickey Mouse, doesn't it?" It has presented a very visible issue. To me, it is not only a question of building the technology but of providing that technology for a particular industrial sector—I would urge that it be the creative sector—to be able to showcase what it can do and, by the way, what we can do as a country. We are an astonishingly powerful creative country and human data is vital for this. If we can have something that can showcase that and support our creative industries, I would back it 100%.

Q126 **The Chair:** Before we leave the question of transparency, concerns have been raised with me and members of the committee about the issues

around X, xAI and Grok. It goes back to this question about being ready to reveal what training data has been used by AI models. Can you give assurances for Google, Roxanne, that you know exactly what material goes into the training of your models, especially those offered to the public sector?

Roxanne Carter: When it comes to training the models, we need, as I said, to train on a huge amount of data. It is not just that you want high-quality data; you need the good, the bad and, in some cases, the ugly as well. For example, if you are training the model to be safe, you need to show it what is harmful. If you are training a model to tackle spam, it needs to be able to analyse spam. But you need to have safeguards so that it learns from or analyses this information but does not seek to produce that content. We take that commitment incredibly seriously. This is not the sort of content that we want to see on our platform. When designing our tools, we have invested in safety right from the get-go. We have very strict policies, but we also have guard-rails to ensure that problematic output such as violent, offensive or sexually explicit content is not shared and generated.

Q127 **The Chair:** If your models range over the bad as well as the good, as you have touched on, how is the content of your training data detailed and listed out? Do you share that information with the UK AI Security Institute?

Roxanne Carter: I know that we publish AI model cards, but I would need to go back and check who we share that with. I will report back to the committee.

The Chair: That would be helpful.

Q128 **Baroness Healy of Primrose Hill:** The committee is very interested in the question of licensing, and that is what I want to return to now. We have heard that one of the main obstacles to the development of a licensing market is a lack of willingness to engage from AI companies. What is your view on that? Is it a fair comment, or is it because you are waiting to see what the outcome of the Government's consultation is?

Roxanne Carter: We have a long track record of doing licensing deals and seeking commercial partnerships with the UK creative industries. With YouTube, we license for use on the platform; we license directly with partners, but also with PRS; we have deals with LyricFind on Search; and for News Showcase, we have over 280 publications that we cover through that scaled solution. We license for content. This comes to what we are licensing for and what we were discussing at the beginning. We do not believe that we need a licence to train, but we are doing licences for access and we are seeing that more and more.

We have done deals. We have just rolled out an AI pilot with news organisations, which the *Guardian* is part of. We are keen to roll it out to more news publishers over the coming months. We have also done deals, for example, with Bloomsbury and Imperial, as we want to make sure that we have specific UK content included in the training of AI

models. These deals are being done. Again, having a text and data-mining exception gives clarity over what we are paying for, and I think that would turbocharge more of these commercial deals being done.

Q129 **Baroness Healy of Primrose Hill:** Mr Gadney, I would be interested in your view as you come from a different perspective in terms of smaller AI companies and your relationship with the creative sector.

Guy Gadney: I first started licensing creative work for digital media in the mid-90s when, as I mentioned, I was head of digital for Penguin Books. In one specific case, we wanted to license 10, I think, 30-second music clips for a CD-ROM database for music. I believe we were quoted £250,000 at that moment in time, which, if I extrapolate today, is around £600,000 for 10 30-second clips. In essence, that was a no to the licence that we sought. More recently, we developed a video game adaptation of John Wyndham's book *The Kraken Wakes*. We licensed the content for that as an adaptation and went through the appropriate channels. It took over a year to secure the licence in that instance.

We have been tracking this for quite some time, fairly obviously. To my mind, there are two sides to the coin: one is the large-scale licensing that your question to Google implies; the other is a licence for use in creative and specific instances. The creative industries look after the creators—remember, we are very creator-centric; we want to see creators being remunerated appropriately and beneficially. The industries, labels, publishers and agents that exist to monetise the works have not done a good job in keeping up with technological innovation.

It is their job to innovate the ways in which they can benefit the creators they represent. I would not advocate them preferring protection and blocking rather than engagement and experimentation. Historically, technologically, that has not worked for the creative sectors nor the creators. As we move into what is now largely a so-called creator economy—I note that research says one in four people in the UK identifies as a creator, which is a significant movement—it is very important that the creative industries and the creators respond to this and engage with this process.

I note that our international AI competitors are on the track already. They are in the stadium and running laps. It feels in the current process of this debate like this country cannot decide whether to enter the stadium at all. If AI is seen as a productivity growth engine, we need leadership to be more decisive, pragmatic and creative in its thinking, seeking to grow and support the creative sectors as they innovate into this sector.

Q130 **Baroness Healy of Primrose Hill:** Following on from that, what specific steps should the Government and regulators take to help create a coherent and sustainable licensing environment? I am also interested in your view on the Government's proposed creative content exchange. Is that worth looking at?

Guy Gadney: We engaged at the beginning under the umbrella of CoSTAR, the government-funded research initiative, which has pulled together a proposal called ACCCT which is looking at transparency. It stands for access, control, consent, compensation and transparency technology framework, which, as you are referencing, is blockchain-enabled. It supports the sort of flexibility, fluidity and speed at which these sorts of licensing agreements need to happen.

I note also that the very nature of content, as we have talked a lot about today, has so far been quite monolithic. There is a work that is licensed for a particular purpose. The nature of content at the moment is much faster than that, much more playful. There is a lot of modding, of taking one piece of content and moving it into another at light speed. For that particular piece, we engaged with that organisation, which was a cross-sector and technological proposal. This has been tabled with DCMS since Q2 last year and could be developed, implemented and leveraged by the AI companies that you refer to. My proposal would be down those tracks, to have something that we can launch test in the market.

Q131 **Baroness Healy of Primrose Hill:** Ms Carter, what role do you think the Government and regulators should take in order to increase the access for AI companies?

Roxanne Carter: Here, I would like to take a step back and say the UK is a hugely significant market for us. It has been home for Google for 20 years. We have just announced a £5 billion investment that will span over two years, £1 billion of which goes to create a new data centre. But having an enabling copyright framework is an important factor in this. For us, having a text and data-mining exception that builds on the non-commercial one that already exists in the UK to a commercial one is quite an important feature that we would like to see adopted by Parliament.

When it comes to the creative content exchange, these sorts of mechanisms can be helpful. They can be very useful in terms of pooling together some of the smaller creators. Again, we would like to see more details. If you ask an entity to map out and try and put a price tag on every single piece of content on the internet, that is a challenge. But if you are pulling this together to reach voluntary agreements so that you get the smaller players to come to the market as well, that could be very useful.

Q132 **The Chair:** Given what you have just mentioned, Ms Carter, about the investment, how important is the investment in the UK that you mentioned, the £5 billion figure? How important are different factors in that decision? Obviously we want to see the continual development of high-quality human-created content, but we have the UK copyright framework. What were the factors that governed or shaped your decisions about where to invest in and develop AI? You mentioned a data centre, for instance. What are the factors that balance out there?

Roxanne Carter: I appreciate there will be many different considerations, but I come back to the fact that the copyright framework is an important factor. I want to be at pains about the importance of getting the balance right. I mentioned at the beginning that the creative industries are incredibly important to Google, and we see them as a partner. We want to make sure that we get that balance of ensuring they can continue to thrive and flourish, as well as allowing AI innovation to take off in the UK. Getting that balanced copyright framework is a top consideration for us.

Q133 **The Chair:** I read out to you the comments that were made to us by witnesses that described Google as a "bad actor", and that talked about you taking unfair advantages in the market. It is a curious thing for you to say how much you value the creative sector at the same time as you are viewed in that way, in taking an unfair advantage in your search and crawler activity and not paying for training data. How do you answer those comments? It is great to talk about partnership, but to the creative sector, there does not look to be much partnership in that sort of behaviour.

Roxanne Carter: If I may, I think the comment of "bad actor" came from the representative of Cloudflare; I may have this correct.

The Chair: No, it was not.

Roxanne Carter: Okay. I do not believe that we are being a bad actor. Like I said, we partner with the industry. We are very keen to develop and grow AI tools. But at the same time, the long tail of human creativity on the web matters, so that is something we are constantly striving to deliver on. Like we discussed, we are doing deals, and we have done pilots. Those will be rolled out and we continue to do more deals. I disagree with the framing of Google as a bad actor.

Q134 **The Chair:** I think it is a combination of the not paying, which we are juggling with all the questions we have asked you on licensing and trying to develop a licensing market, but everything hinges in many ways on this remuneration question and being prepared to pay for use of data for training. It was very strongly put to us that if Google will not pay—and you are a big player in this market—why should other people pay? There is a leadership role there. I just reflect back to you that it comes across a bit hollow to talk about partnership with the creative sector if you are one of the players who take advantage and will not pay. I will leave you with that point, unless you want to make any more reflections on it.

Roxanne Carter: Just to reiterate, it is not a question that we will not pay—we need the certainty as to what we are paying for. Those deals are already being done, and we will continue to do those deals.

Guy Gadney: I applaud that Roxanne and Google value the creative industries. My comment is that they are not valued enough. Putting a value on them needs to be quantified, not just said.

Secondly, the payment terms we have started to see coming out of settlements are one-offs to creators, for example through the Anthropic case. That is unacceptable in a traditional licensing model where trailing revenues—you talk about the long tail—are not accepted as part of the deal. Any traditional licensing model usually has some form of revenue share or points associated with it. The one-off deal does not work, having spoken to writers. In my opinion, something needs to be done around that long tail that we talk about.

Thirdly, we have seen a huge shift in revenues. We must acknowledge the shift in revenues from traditional media that we talk about. Revenue has shifted over 60% from linear to digital platforms over the last 10 years. Advertising income has shifted over 77% on to Google, Meta and other advertising platforms. That has had an enormous impact, as we know, on our national linear advertising networks. Those networks commission content. They do not look at the long tail; they look at the short tail. They look at commissioning and funding content that we talk about as copyright material.

Forget copyright for a second, because I believe it can be a cloak towards what we are actually talking about, which is money. In a very un-British way, let us start talking about money. Follow the money, look at where the money is going, who has it, and how much of that money is getting through to creators themselves. Those are the people who are creating the content. That is a healthy ecosystem; that, as a nation, is what we are good at.

The trouble with supporting the long tail is that it implies the content already has to exist. Who is funding the creation of that content? We can create our content at an incredibly low cost, which is great, and there is an argument that we can democratise that content. That does not mean that people should not eat or have an unsustainable lifestyle, simply as they have to then create this and rely on a long tail, which is substantially less than the existing model. So, as we are looking into this sort of model, I urge that we look at the money and at how we can support the creation of the content, as well as the long-tail remuneration for it.

The Chair: Thank you very much. We started with a question of payment or revenue share, and we ended on the same topic: let us look at the money. Thank you both very much indeed for coming today.