British Copyright Council: House of Lords Governance of artificial intelligence (AI) Inquiry Response

The British Copyright Council (BCC) represents those who create, hold interests, or manage rights in literary, dramatic, musical, and artistic works. The following response has been developed with our membership which include professional associations, industry bodies and trade unions which collectively represents the voices of over 500,000 creators, spanning the creative industries.

These right holders include many individual freelancers, sole traders, and SMEs, as well as larger corporations within the creative and cultural industries. Our members also include collecting societies which represent right holders, and which provide licensed access to works of creativity. A list of our members can be found here.

Many BCC members are creators who increasingly work with AI technologies as both assistive and generative tools linked to the works they create. On the other hand, many creators are extremely concerned with good reason, that AI-outputs are and will be used instead of human-authored work. As such, transparency over how creative works can be ingested and adapted throughout this process will be increasingly important and IP licensing safeguards will remain vital to protect against the unfair use and devaluation of copyright protected work. This can be accomplished by respecting existing UK copyright and related rights laws.

The Committee is seeking evidence on the following questions (there is no requirement to answer all questions in your submission):

Capabilities and trends

1. How will large language models develop over the next three years? Given the inherent uncertainty of forecasts in this area, what can be done to improve understanding of and confidence in future trajectories?

N/A

BRITISH COPYRIGHT COUNCIL

2. What are the greatest opportunities and risks over the next three years? How should we think about risk in this context?

As matters currently stand, open-source foundational models pose a huge risk to the UK's creative industries, especially if they remain unregulated, since they provide an opportunity for both data laundering and unauthorised copyright use- thereby causing mass-scale copyright infringement.

There are demonstrated bad actors already doing this. LLMs, generative AI, and opensource models will likely continue to pose specific challenges for regulators trying to determine legal responsibility for AI outcomes from an infringement perspective. Since these models are often extremely complex, it can be difficult to understand how they make decisions or generate output and what data has been used to help the model reach this decision. This makes it challenging to identify specific causes or cases of copyright and related rights infringement and to determine who is responsible for them.

Furthermore, these models are often developed by multiple individuals and organisations, and across jurisdictions, which creates further difficulties in determining who is responsible for negative outcomes that may arise. The use of open-source models in particular can further complicate this issue, as it may not be clear who has contributed to the development of the model itself.

Since these models are trained on large datasets that include a wide range of content, which may or may not be under the protection of copyright and related rights, this makes it potentially difficult to determine who is responsible for any specific data point, decision, or inclusion of infringing copyright material. Determining ownership of the training data and whether it has been lawfully obtained can be challenging, particularly when the data has been collected from multiple sources. The output generated by foundation models itself may also contain copyright content, such as text or images. Additionally, the use of transfer learning techniques can make it challenging to understand how a particular model has been influenced by prior training on other datasets which may have infringed on copyright and related rights.

To address these challenges, regulators will need to develop new approaches for determining legal responsibility and agreed standards for AI outcomes that include looking at the originating data/creative works at their source– addressing accountability, traceability and transparency matters. This will likely require collaboration between different stakeholders, including the creative sector, industry, academia, and civil society, to ensure that these issues are addressed in a comprehensive and effective manner.

Having agreed industry standards for LLMs will play a crucial role in applying a transparent and certain approach across sectors. They will be invaluable and should be agreed in consultation between all stakeholders.

BRITISH COPYRIGHT COUNCIL

3. How adequately does the AI White Paper (alongside other Government policy) deal with large language models? Is a tailored regulatory approach needed? What are the implications of open-source models proliferating?

The white paper currently outlines 5 clear principles that these regulators should consider to best facilitate the safe and innovative use of AI in the industries they monitor. The principles are:

- safety, security and robustness: applications of AI should function in a secure, safe and robust way where risks are carefully managed
- transparency and explain-ability: organisations developing and deploying AI should be able to communicate when and how it is used and explain a system's decision-making process in an appropriate level of detail that matches the risks posed by the use of AI
- fairness: AI should be used in a way which complies with the UK's existing laws, for example the Equality Act 2010 or UK GDPR, and must not discriminate against individuals or create unfair commercial outcomes
- accountability and governance: measures are needed to ensure there is appropriate oversight of the way AI is being used and clear accountability for the outcomes
- contestability and redress: people need to have clear routes to dispute harmful outcomes or decisions generated by AI Instead of specific legislation

On a general note, these principles are welcome but must be explored and strengthened further to be fit for purpose.

For example, disclosing and clearly labelling the use of AI regarding both final outputs and ways of working is a necessary baseline. However, additional measures would be required to ensure that there is suitable transparency in the input stage regarding the use of copyright works to help train generative AI algorithms. However, while labelling within the context of supporting transparency for users of outputs may be helpful for those responsible for managing potential regulatory liabilities for the use of AI, it is not going to be sufficient, or enforceable (at present), to address all cases. For that reason, any requirements need to be both well thought out and practical to implement across different types of AI tools. For example, disclosure that AI is being used to evaluate credit or medical records would look different than disclosure that is required for an image or bodies of text produced by generative AI.

Furthermore, transparently disclosing the use of AI would help mitigate consumer rights related issues we are seeing beginning to emerge. For example, the dissemination of deepfakes and fake news formed from generative AI outputs which can pose a real threat to society.

BRITISH COPYRIGHT COUNCIL

4. Do the UK's regulators have sufficient expertise and resources to respond to large language models?[5] If not, what should be done to address this?

There is a clear role for regulation.

However, the UK faces a challenge when it comes to regulating the infringement of copyright works by many large language models since the IPO is not a regulator in its own right despite playing a critical role in promoting and ensuring the copyright law is recognised and applied effectively across all parts of the economy.

For regulation to be effective, regulators must be linked to recognition of the role played by the IPO, as well as the key features of the UK copyright regime, and must be suitably resourced and empowered to take on the oversight and enforcement measures which are inherent to the development of ethical AI. This will also require the lateral interaction between various regulators in particular the ICO and the CMA as well as the IPO whose regulatory role at the moment is very limited.

In principle, the cross-sectoral principles outlined provide a useful starting point for the formation of a central regulatory coordinating function. However, we see clear challenges in implementation which will need to be addressed to avoid eroding the UK's copyright regime and causing irreparable harm to the economic value of our creative industries. In particular, the principles as currently outlined do not adequately address application of the laws of copyright and related rights or data protection regulation applicable for inputs and the intended purposes of certain AI developments, such as generative AI.

Furthermore, Part Three, paragraph 34 of the current White Paper, states that their proposed regulatory framework does not seek to address the balancing of the rights of content producers and AI developers. We strongly urge for this to decision to be re-examined. Any reasonable framework needs to consider the rights of creators and rights holders – whose works are ingested to form machine learning datasets used to train AI programs – as maintaining transparency measures is a keyway to maintain fairness, economic viability, and legal accountability for all. Transparency over machine readable systems which fully consider the issue of licensing consents, alongside traceability, is critical to ensuring the possibility of a well-functioning market.

Any Regulator responsible for oversight of any AI applications linked to different sectors of the economy must have a duty to highlight and support compliance with both the laws of copyright and related rights and data protection laws as a fundamental principle for application of any other sector specific regulatory provisions addressing issues of liability. We are conscious that a nuanced approach may be required to fully support innovation across sectors. However, copyright, related rights, IP and data protection are areas of law that will be relevant for all

sectors and should be used as a robust baseline while sector-specific liability issues may need further guidance from sector-specific regulators for finance, defence, medicine, consumer rights and so on.

This inputs issue in particular cannot be ignored and is key to having a comprehensive set of principles which is fit for purpose. Furthermore, there needs to be a meaningful backstop where AI firms might not be good actors/compliant, to foster market access, which is predicated on compliance. Ensuring that regulation considers the works of creators and rights holders is critical if we are to develop a regulatory landscape which facilitates the development of AI in a transparent and accountable manner. This might include:

- Using authorised data sources: When training AI models, using data from authorised sources, including through licensing where appropriate and is permission-based, will be crucial to avoid infringing on copyright and related rights. Additionally, care should be taken to ensure that data is used only for the intended permitted purpose.
- Implementing internal access controls for AI developers: to limit who can access and use copyright materials which would also serve to avoid inadvertent infringement and help prevent unauthorised use of works in AI development. **It should not be the responsibility of rights holders to "opt out". Rather it should be the responsibility of anyone who wants to use protected data to make sure they are authorised to use this data.** The nature of scraping for massive datasets by which machines are subsequently trained means that machines initially learn from the ingested dataset and do not unlearn. Rather it should be the responsibility of anyone who wants to use protected data to make sure they are authorised to use this data. The assumption should be that authorization is not granted unless such authorization is explicit. Furthermore, rights holders should be able to rely on copyright and related rights protections currently in place in order to avoid implementing measures that could otherwise hurt their ability to publicly display their work.
- Obtaining licenses: Obtaining licenses for copyright materials would also help ensure that works that were used to train application of an AI algorithm or program remain protected in their own rights and rights holders are remunerated across the commercial stages of the model. Any licensing is based on the concept of right holders' choice, it is a decision for the creator whether they want to allow the use of their work for AI purposes including the pre-training of AI foundation models. Transparency will enable clear assessment of the nature of this use.
- Monitoring and logging usage: Monitoring usage and tracking copyright works accessed while ensuring that these logs remain transparent and accessible would help identify and address potential violations of copyright and related rights. This could involve using tools to detect unauthorised use of copyright materials or monitoring user behaviour to ensure compliance with licensing agreements.

By taking such steps, in coordination and partnership with rights holders and creators, AI developers can ensure that they are developing AI applications in a responsible, ethical,

transparent and legal manner while also safeguarding the future of the industries upon which they depend for inputs, thereby ensuring future high-quality outputs.

We believe that express reference to the vital role that can and should be played by copyright and related rights in developing the transparency, fairness, accountability and contestability and redress principles should be explicit. In addition, the government's approach would benefit from the inclusion of principles which explicitly relate to the input, accountability, and transparency of training data. This should include obligations to make sure that data is sourced legally and that rights holders have transparency and traceability mechanisms that allow them to freely choose to participate- or not- in the value being created from their works.

5. What are the non-regulatory and regulatory options to address risks and capitalise on opportunities?

The case for regulation

As noted in our response to question 4, government can and should play a role in creating a legal and regulatory framework that supports innovation in AI development while preserving copyright. Regulators could take measures, in collaboration with rightsholders and creators, to develop guidelines and standards for AI development within sectors of the economy that include specific requirements for compliance with copyright law. Regulators could also be empowered to monitor AI development and actively take enforcement action themselves against AI developers who infringe on copyright. Government and regulators hold the responsibility of ensuring accountability of AI development.

We welcome the stated intention that a new central coordinating function would also be charged with promoting further collaboration between regulators. The IPO for copyright and related rights and the ICO for data protection regulation in particular must have key roles here to protect and enable effective ongoing application of existing frameworks. However, for such a function to be fit for purpose, it will need to be appropriately resourced. We recommend setting clear guidelines for all regulators, and where required introduce cross-sectoral statutory duties, to ensure sectors such as the creative industry do not become siloed and fragmented between regulators such as the ICO and CMA for example; too many regulators for one sector will only lead to confusion rather than improving transparency. More consideration should also be given to addressing how the proposed framework could address the "inputs" question while incorporating meaningful enforcement mechanisms.

However, we believe more work needs to be done to ensure that the UK's copyright regime is not inadvertently undermined and as a result impacting our creative industries contribution to the UK economy. Doing so will future-proof the AI regulation framework and would foster more trust and cross-sectoral collaboration. The failure to do so would not only cause harm to

the known strengths of the UK's IP industries but would also undermine the ability to promote innovation.

In particular, a clear and uniform structure for IP compliance and data protection compliance would provide support for all regulators looking to the IPO and the ICO for guidance in these areas, as they look to develop sector specific liability. Such a duty could provide regulators with a clear mandate to act against AI-related copyright and related rights infringement and ensure that the necessary measures are in place to prevent and address such infringement. Duties on regulators could include an explicit requirement to develop guidelines and standards for AI development that preserves copyright laws and the needs of copyright holders.

UK regulators could also be required to index and monitor AI development and afforded the power and authority to take enforcement action against AI developers who infringe on copyright. Additionally, given the impact on data protection we strongly suggest a specific statutory duty on the ICO to recognise the protections which are relevant for text and data which attracts rights of copyright and related rights.

We also acknowledge there are also potential challenges with implementing a statutory duty on regulators especially regarding resource constraints which would make it difficult to effectively monitor and enforce copyright and related rights laws in the context of AI development. Such monitoring must remain the responsibility of government and the IPO. Nevertheless, providing appropriate governmental safeguards in relation to AI, such as introducing the suggested requirement of recognition for all regulators would certainly help to clarify and strengthen regulators' mandates. Failing to do so leads to a danger of creating clashing and confusing guidelines which inadvertently allow for infringement in some sectors to slip through the net.

Exploring the possibility of government certification and licensing requirements that can be revoked if an AI system or product is found to have been developed in a manner inconsistent with these principles may also be an avenue worth exploring. Revoking permissions in cases where there are bad actors or non-compliance would also strengthen the regulators' role, otherwise monitoring and evaluation are pointless.

Non-regulatory tools for trustworthy AI

We will limit our answer to the copyright and related rights perspective and would like to take the opportunity to note that there are several market-based solutions for licensing which are currently in place and many more are being developed by rights holders. For instance, those provided by big publishing houses such as Elsevier, image libraries and agencies, and collective management organisations (CMOs). It is crucial that government intervention does not inadvertently result in undermining what is still a relatively nascent and innovative market for rights holders.

From a non-regulatory perspective, we also encourage government to work directly with industry, to develop best practices and guidelines for future foundational AI development that consider copyright laws and the needs of copyright holders. The British Copyright Council and our members would be especially well placed to take this forward.

Ultimately, the most helpful non-regulatory tools for embedding AI regulation principles into existing business processes will likely depend on the specific context and needs of each organisation and a combination of different tools and approaches may be necessary to achieve the desired outcomes. However, the following non-regulatory tools could be helpful in this regard if they are developed in a manner which is fully consistent with the UK's current copyright framework:

- Standards and guidelines: Standards and consistent copyright and related rights adherent guidelines could help provide clear and consistent guidance on how to design, develop, and deploy AI systems in a trustworthy manner which provide for compliance with existing copyright law.
- Best practices: Best practices could be shared to provide practical guidance on matters such as securing suitable licensing or working with other permissions based systems; logging and auditing inputted works; and labelling relevant AI-generated works as such. This would also ensure that AI learning tools are not built upon unauthorised content or material that rightsholder wish excluded from such uses. However, promoting best practice does not replace the need for clear regulation on transparency and auditability.
- Training and education: Training and education on matters relating to AI, copyright and related rights, in the form of training and guidance, could help organisations to build the knowledge and skills necessary to pursue innovation while avoiding infringing on the rights of creators and rights holder. To stress, this is not a zero-sum game.

5.a How would such options work in practice and what are the barriers to implementing them?

To be completed following CTWG discussion

5.b At what stage of the AI life cycle will interventions be most effective?

To be completed following CTWG discussion

BRITISH COPYRIGHT COUNCIL

5.c How can the risk of unintended consequences be addressed?

To be completed following CTWG discussion

International context

6. How does the UK's approach compare with that of other jurisdictions, notably the EU, US and China?

At this stage, we are not aware of any existing organisations that deliver a central regulatory coordination function as is proposed by the AI White Paper. However, other jurisdictions including the EU are in the process of building this out. The OECD Artificial Intelligence Policy Observatory, which facilitates dialogue and provides multidisciplinary, evidence-based policy analysis and data on AI's areas of impact, could be helpful in providing policy analysis. More specifically, taking the lead on safeguarding the UK's IP framework, which the UK has successful developed over decades, would be astute. It is key that there is harmonisation with other jurisdictions so that UK developed technology can be exported to other international markets. A specific regulator equipped with the appropriate skills might be required to ensure holistic approach which will require collaboration.

Avoiding overlapping, duplicative, or contradictory guidance on AI issued by different regulators will require effective coordination and collaboration between regulators at national and international levels- a central coordinating body could help with this. Strengthening the role of the UK IPO and resourcing it to take on new coordinating and enforcement responsibilities with respect to AI and copyright and related rights would also be of benefit to avoid the creation of overlapping, duplicative or contradictory guidance on AI. This would support the role of the IP and AI Minister appointed by the government.

TO BE DISCUSSED WITH THE CTWG

EU

Japan

United States

6.a To what extent does wider strategic international competition affect the way large language models should be regulated?

To be completed following CTWG discussion

6.b What is the likelihood of regulatory divergence? What would be its consequences?

To be completed following CTWG discussion