Artificial intelligence (AI) and intellectual property (IP)

An open letter to the Intellectual Property Office (IPO) in response to the UK government's consultation on proposed reforms to copyrights and patents

7 January 2022

Contact: policy@theodi.org

Background

On 26 October 2021 the Intellectual Property Office (IPO) published an [open consultation on Artificial Intelligence and Intellectual Property: copyright and patents](#) featuring proposals for reform to the intellectual property (IP) system and an accompanying [impact paper](#). The consultation closes on 7 January 2022.

The IPO has the ambition of making reforms to the UK IP system that will incentivise research, development and deployment of artificial intelligence (AI) technology. The consultation questions and impact assessment recognise the importance of data availability for the development of a world-leading AI sector in the UK, and puts forward proposals that would improve data availability for businesses and other organisations.

The [Open Data Institute (ODI)](#) welcomes these aims. But we believe that it is important that the IP regime takes a strategic and holistic approach to addressing market imbalances in the AI ecosystem for the best economic and social outcomes for everyone.

The data value chain and the role of IP for AI

At the ODI, we want those who steward data and those who create information from this data to act in ways that lead to the best economic and social outcomes for everyone. Our [Theory of Change](#) contrasts this positive data future against two negative alternatives: a *data-hoarding* future, in which inappropriate business or funding models lead to data monopolies; and a *data-fearing* future, in which inappropriate use of data leads to mistrust in data sharing. Both a data-fearing future and a data-hoarding future jeopardise opportunities for data use for the best outcomes.

Access to data is vital for developing and training AI systems. But in the UK this is at risk of being impacted by market imbalances that could lead to data-fearing or data-hoarding futures. [Our research](#)[1] has identified a pattern where organisations with access to large siloes of high-quality data are better able to train high-performing AI systems with it; these high-performing AI systems are more likely to be more widely adopted; and the wide usage of these high-performing AI systems generates more data for the data siloes, in a feedback loop that tends towards monopoly. The [Furman Review (2019)](#) argues that the challenges to effective competition in digital markets are amplified by 'network-based and data-driven platform business models [that] tend to tip markets towards a single winner', and that data openness must be a key tool for the new [Digital Markets Unit](#) for promoting competition.[2] The tendency to monopoly also narrows participation in innovation with data or in innovation with AI, which can reduce the scope of this innovation and communities' experiences of it. At the ODI

---

[1] [The role of data in AI business models [report] (The ODI, 2018)](#)
[2] [Unlocking digital competition: Report of the Digital Competition Expert Panel (HMT, 2019)](#)

we think it is important that interventions for supporting data availability ameliorate, rather than replicate, these market imbalances to avoid data-hoarding or data-fearing futures that will have negative consequences for the future of the AI ecosystem.

Our Theory of Change identifies three stages along the data value chain where policies or practices by governments, businesses, civil society or regulators can influence the kinds of data futures that develop – the IP system affects all three. There is the **stewarding stage**, where data is collected, maintained and shared; then there is the **information-creating stage**, where information is created from data in the form of products and services, analyses and insights, or stories and visualisations; and finally there is the **decision-making stage**, where decisions are made about the outcomes of the information-creating stage, alongside multiple sources and drawing on experience and understanding. We believe that there is both a need and an opportunity for reforms to the IP system to strengthen data availability across the economy, supporting the development and deployment of AI systems at each of these stages of the data value chain:

- stewarding data (reflected in Section 1 of the consultation on copyright and database rights);
- creating information (reflected in Section 2 of the consultation on text and data mining); and
- decision-making (reflected in Section 3 of the consultation on patents).

Consultation section 1: copyright and database rights

Computer-generated works without a human author are currently copyright-protected in the UK for 50 years. The IPO is asking whether to keep these protections, reduce their length or scope, or remove them entirely. Another type of protection is database rights: similar but distinct rights that give protection to the contents of a database for 15 years, including those databases produced by AI systems. Examples of these databases include, but are not limited to: synthetic datasets, which are frequently used for privacy protecting raw data; datasets created through feature detection on satellite imagery; and databases of artificially generated faces. These kinds of datasets are likely to become more prevalent with uptake of Privacy Enhancing Technologies (PETs) such as anonymisation techniques.[3]

An IP regime typically influences the stewarding stage of the data value chain, as a mechanism by which data stewards can control access to or use of data that they steward along the data spectrum from 'closed' to 'open'. This is important because firms often lack incentives to share data that they have stewarded: ODI research[4] in 2018 found that many AI companies do not believe that they will benefit from sharing or opening their data to others. Studies show that data sharing typically creates more value for the wider economy than for individual data stewards: OECD research in 2019 suggests that 'data access and sharing can increase the value of data to holders (direct impact), but it can help create 10 to 20 times more value for

---

[3] Anonymisation and synthetic data: towards trustworthy data (The ODI, 2019)
[4] The role of data in AI business models [report] (The ODI, 2018)

data users (indirect impact), and 20 to 50 times more value for the wider economy (induced impact)'.[5]

Database rights provide a mechanism through which organisations can require payment for others to use databases they have stewarded, including when those databases have been generated by an AI system. This creates revenue streams and therefore incentives that would not otherwise exist for certain types of data-sharing. For example, database rights can help to facilitate data-sharing through licensing, supporting data availability and consequently innovation using data (such as the development of AI systems). We believe the IPO is right to recognise and support this dynamic, but we think the limitations should also be considered: *sui generis* database rights only exist in a small number of jurisdictions, and a 2017 public consultation carried out by the European Commission found mixed reports of their effectiveness [6]; legal and economic scholarship also challenges the viability of exclusive property rights for data[7].

Although database rights can help to create incentives for some data stewardship activities, they don't provide incentives for some important types of data-sharing (such as opening-up datasets) and can also be used in ways that end up reducing data flows. For example, new participants in a sector may have little option but to rely on existing databases controlled by the early movers in the sector, leading to a 'winner-takes-all' model[8] whereby an incumbent data provider can command an overwhelming competitive advantage because new efforts by others at collecting data might not always be efficient. This acts as a barrier to entry or participation in the market, which in turn impacts future data availability: the Creative Commons Copyright Platform AI Working Group has argued that database rights are a potential harm to the development of AI, and should be repealed.[9]

We believe that any intervention around database rights for databases generated by AI systems to increase data availability must factor in some key considerations. Firstly, when databases generated by AI systems are shared, it is important that those databases are properly assured for trustworthiness. This assurance could include, for example: verifying the quality and completeness of the database; establishing rights and consent for the database to be shared or used in new ways; and providing clear and transparent rules on licensing for databases produced by AI systems. Secondly, we think it is important that the use of databases generated by AI systems is accompanied by strong data literacy for critical thinking about how that data has been generated and stewarded, and the questions that are asked of it, the models that are applied to it, or the use-cases to which the database is applied. Interventions to support data literacy and data assurance may not fall directly within the remit of the IPO as it has traditionally

---

[5] Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies (OECD iLibrary, 2019)

[6] Synopsis report of the public consultation on the evaluation of Directive 96/9/EC on the legal protection of databases | Shaping Europe's digital future (European Commission, 2018)

[7] The argument against property rights in data: ensuring access to data though non-exclusive access rights (Open Future, 2021)

[8] The role of data in AI business models [report] (The ODI, 2018)

[9] Wardeh, M. M. (2021). Text and Data Mining. *Creative Commons Copyright Platform AI Working Group*

operated, but we think that it is critical that these considerations be factored in alongside the reforms proposed in this consultation.

Consultation section 2: Text and Data Mining (TDM)

Text and data mining (TDM) is the use of automated computational techniques to analyse large amounts of information to identify patterns, trends and other useful information. Current UK IP law permits the use of TDM on any copyright work for non-commercial research purposes. The IPO is considering a range of different options to make TDM easier to carry out, from improving licensing to permitting the use of TDM on copyright work for commercial research purposes too.

While interventions for supporting data availability for development of AI systems can add value to the wider economy, these should not be done in a way that undermines longer-term trust in data practices. A 2021 report[10] by Frontier Economics for the ODI found robust evidence that trust is a key determinant of data sharing: if data is shared in ways that result in harm, this could reduce trust in data use and therefore levels of data sharing overall.

In ODI's theory of change, an AI system is typically at the information-creating stage of the data value chain, allowing sophisticated analyses and insights to be generated from data. But analyses created with AI systems can result in unfair outcomes that create or reinforce harms. The controversy in 2020 over the use of an algorithm to assign A-level students' grades shows that public trust in the use of data and AI systems can be damaged when the development or deployment of an AI system is perceived to be biased or unfair.

Public datasets are not always representative of diverse communities and so the indiscriminate use of TDM to train algorithms may lead to the replication of existing biases, the codification of taxonomies or bodies of knowledge[11] that are limited in their usefulness, and the creation of blind spots about different kinds of questions or analyses. The Creative Commons Copyright Platform AI Working Group has argued that one way of avoiding these pitfalls is by ensuring the maximum volume and widest diversity of content is available for training purposes.[12] We believe that strengthening ecosystems of data institutions (ie organisations that steward data on behalf of others, often towards public, educational or charitable aims) can improve access to a wider range of relevant datasets for TDM while providing appropriate governance for its use. Data assurance and data literacy will also play a key role in identifying and mitigating risks of bias in TDM.[13]

We also think it is important that diverse organisations across a wide range of sectors are supported to benefit from any relaxation of the law around TDM. This might include supporting data capabilities and infrastructure in sectors and domains that are comparatively behind in adoption of AI systems. Interventions to support diversity in datasets for TDM and diverse

---

[10] The economic impact of trust in data ecosystems – Frontier Economics for the ODI [report] (The ODI, 2021)

[11] The UK National Data Strategy 2020: data as infrastructure, data for infrastructure (The ODI, 2020)

[12] Wardeh, M. M. (2021). Text and Data Mining. *Creative Commons Copyright Platform AI Working Group*

[13] The ODI's written evidence for APPG on Artificial Intelligence (The ODI, 2021)

participation in TDM may not fall directly within the remit of the IPO as it has traditionally operated, but we think that it is critical that these considerations be factored in alongside the reforms proposed in this consultation.

Consultation section 3: Patents

UK patent law allows humans that use AI to devise inventions to be named as inventor in most cases, but does not currently recognise AI systems as inventors. However the IPO has already received two patent applications which name an AI system as inventor. The IPO is considering a range of options to better reflect AI-devised inventions in the patent system, from expanding the definition of 'inventor' to include humans responsible for an AI system that devises inventions, to allowing AI systems to be identified as inventors.

In the ODI's theory of change, the final decision-making stage of the data value chain is where people and organisations decide what to do with the outcomes of the information-creating stage - such as deciding what to do with insights or analyses generated by AI systems. IP regimes can also influence the final decision-making stage of the data value chain, by creating incentives for these insights to be developed, used or shared in ways that don't lead to data-fearing or data-hoarding futures. We believe that reforms to the patent system should not reinforce existing concentrations of market power and access to data. For example, advances in AI could result in large volumes of patents being held by a small number of dominant players that have access to the most-developed AI technology derived from the largest data siloes for training AI systems.

We think it's important that the IP regime doesn't disproportionately inhibit the development and use of open models in AI, which encourage innovation in the development of AI systems by allowing greater scrutiny and rigour. For example, during the Covid-19 pandemic sharing the models underpinning different aspects of emergency-response decisions supported a wide range of tools and international collaborations, improving the quality of decisions as well as developing wider trust in them, through transparency and reproducibility of science.

Open models can also widen participation in AI innovation by lowering the barriers to participation. This can help ensure that innovation in AI systems reflects the needs and interests of diverse communities, as well as diverse communities bringing fresh perspectives to AI innovation opportunities. Our work on DataPitch has also demonstrated that open models can improve the equitable distribution of the benefits of innovation by enabling access for a broader range of organisations and communities – for example, not limited to capital cities and other traditional 'tech hubs'.[14] [15]

We believe that opening up AI models can lead to similar efficiencies and benefits as opening up data[16], such as saving on internal resources for the organisations developing the models or enabling others to develop customer-orientated services and innovations linked to the models. The Creative Commons Copyright Platform AI Working Group has argued that it is 'unwise to

---

[14] Data Pitch evaluation (London Economics, 2019)
[15] Tech Nation Report 2021
[16] Policy to unlock the economic value of data (The ODI, 2021)

attempt to force the application of an already outdated system [of copyright] to the nascent and uncharted field of AI technology'[17]; we think it's important to take a holistic view of how the IP regime can support the UK's broader AI ecosystem, factoring in the breadth of innovation opportunities and benefits.

Sustainable AI futures

We welcome the ambition of the IPO to strengthen data availability for AI systems, as well as data availability from AI systems, so that the value of this data for economy and society can be realised.

But data availability must be trusted and trustworthy to be sustainable, and to avoid either data-hoarding or data-fearing futures. So we encourage the IPO to think strategically and holistically along the data value chain from stewarding data, to creating information with data, to deciding what to do with that information; and to think strategically and holistically about the role of IP interventions at each of these stages.

We believe that for the government to fulfil its ambitions for the UK to be 'one of the very best places in the world to live with, work with and develop AI'[18], data availability is only part of the equation: consideration must also be given to the assurance of that data, literacy around data use, the role of open models and open innovation in strengthening AI innovation, and the role of data institutions in governance of data for AI.

An IP regime that supports the development of AI systems in an open and trustworthy way will be vital to the delivery of the Government's priorities around national digital and innovation goals as set out in the Innovation Strategy (BEIS), AI Strategy (BEIS), National Data Strategy (DCMS), and for delivery against the ambition of Plan for Growth (HMT) and the Integrated Review (Cabinet Office/FCDO) for the UK to be a global leader in digital technologies and AI.

---

[17] Wardeh, M. M. (2021). AI Generations / Creations. *Creative Commons Copyright Platform AI Working Group*
[18] AI Roadmap (UK AI Council, 2021)