

We urge the UK Government to ensure the UK is a favourable place to develop and use safe AI, by clarifying that public and legally accessed data is available for AI training and analysis in its Code of Practice.

We, the undersigned organisations, welcome the Intellectual Property Office's efforts in taking forward the AI recommendations contained within the Vallance Report. While many other countries have clarified their intellectual property laws to support AI and innovation, the UK has yet to introduce a text and data mining exception to explicitly support knowledge transfer and commercial AI. Given this, the Code of Practice provides a particularly important opportunity to provide clarity and ensure that the UK remains an attractive place to undertake and invest in machine learning.

As rights holders, researchers and innovators, we understand the importance of a well-functioning IP system which strikes an appropriate balance between protecting intellectual property rights and providing the necessary limits and exceptions to those rights, in order to ensure we have the right incentives to create, innovate and develop knowledge. Even without an explicit commercial text and data mining exception, other exceptions and legal doctrines will allow for text and data mining on copyrighted works.

Whilst questions have arisen in the past which consider copyright implications in relation to new technologies, this is the first time that such debate risks entirely halting the development of a new technology.

AI relies on analysing large amounts of data. Large-scale machine learning, in particular, must be trained on vast amounts of data in order to function correctly, safely and without bias. Safety is critical, as highlighted in the Bletchley Declaration. In order to achieve the necessary scale, AI developers need to be able to use the data they have lawful access to, such as data that is made freely available to view on the open web or to which they already have access to by agreement.

Any restriction on the use of such data or disproportionate legal requirements will negatively impact on the development of AI, not only inhibiting the development of large-scale AI in the UK but exacerbating further pre-existing issues caused by unequal access to data.

It will create barriers to entry and raise costs for new entrants.

Unlike other countries, it also would mean that AI model developers would be unable to train their models on publicly available data in the UK without an explicit licence from each rightsholder. In addition to making the UK uncompetitive in AI markets it will disproportionately impact small to medium enterprises, knowledge transfer and hinder open source development of AI.

Importantly, text and data mining techniques are not only used to train AI. Text and data mining techniques are necessary to analyse large volumes of content, often using AI, to detect patterns

and generate insights, without needing to manually read everything. Such analysis is regularly needed across all areas of our society and economy, from healthcare to marketing, climate research to finance.

We believe that in order to support and incentivise researchers and innovators, the UK is best served by a balanced copyright system that encourages the many exciting economic and social opportunities that AI makes possible.

In order that the UK remains competitive in scientific and technology markets, the government should ensure that a Code of Practice:

- Clarifies that access to broad and varied data sets that are publicly available online remain available for analysis, including text and data mining, without the need for licensing.
- Recognises that even without an explicit commercial text and data mining exception, exceptions and limits on copyright law exist that would permit text and data mining for commercial purposes.
- Recognises that the UK operates in an international environment where global norms to support AI are well developed. We observe that countries such as the US, Israel, South Korea, Singapore and Japan have broad fair use doctrines or text and data mining exceptions of differing levels of flexibility aimed at supporting research and technological advancement. As these countries have concluded, we believe that even small differences in clarity in IP regimes can result in big effects on the economy.
- Recognises the broad application of AI across many other sectors of the economy; not least health, the environment, bio-science, agriculture, transport, logistics etc. The Code of Practice meetings have focused on the requests of the creative industries. The report should reflect this by focusing on the outputs of AI systems that are relevant to that sector.
- Supports the Prime Minister's vision that the UK becomes a world leader on safe and responsible AI. The ability to train AI models on broad and varied data sets that are publicly available or legally accessed under agreement will enable the development of safe, ethical and unbiased AI. The code should emphasise the role of government in supporting high-functioning AI by ensuring that all entities are able to develop AI using the necessary scale of data.
- Avoids introducing frictions around using data necessary to develop safe AI. Any measures that discourage the use of broad and varied data sets will have a serious and negative impact on all sectors of the economy using AI. AI is predicated on the three Vs – velocity of processing power as well as volume and veracity (data that is required to ensure the models are accurate). Introducing frictions that hinder use of publicly available or legally accessed data for training therefore prevents the necessary use of data of all types that is required to make sure the models have high levels of predictive accuracy and avoid bias. We must support accurate AI and not hinder it.

In terms of specific features of a Code of Practice, the following should be included:

- Explicit reference to the idea-expression dichotomy, existing exceptions, limitations and the implied licence doctrine that allow the processing of data that a person or organisation already has legal access to. We must not interfere with laws that support the very functioning of the internet.
- Clarifies that access to broad and varied data sets that are publicly available online remain available for analysis, including text and data mining, without the need for licensing.
- Clarification that the code of conduct does not undermine the operativity of s29A (Copies for text and data analysis for non-commercial research.)
- In order to promote the uptake of AI, support and encourage the creative industry sector to develop standardisation of data access agreements, data and schemas.
- The establishment of service level agreements with content providers to swiftly address instances where access to paid for and legally accessed data is erroneously suspended, contrary to the terms of the agreement entered into. This is a particularly acute issue for scientific researchers in universities. The UK should follow international best practice and require that access for data analysis and AI is reinstated within a maximum of 72 hours.
- Support collaboration between AI developers and content creators to mitigate the risk of AI tools from being used to infringe intellectual property rights and promote disinformation.

IP Federation

European Alliance for Research Excellence

The Entrepreneurs Network

Knowledge Rights 21

Creative Commons

Wikimedia UK

Research Libraries UK

Libraries and Archives Copyright Alliance

OpenUK

BSA | The Software Alliance

10th November 2023