

## ARTIFICIAL INTELLIGENCE AND INTELLECTUAL PROPERTY: COPYRIGHT AND PATENTS

1. We write on behalf of the News Media Association (the “NMA”), the voice of UK national, regional and local newspapers in all their print and digital forms - a £4 billion sector read by more than 46 million adults every month in print and online. Our members publish over 900 news media titles - from The Times, The Guardian, the Daily Mail and the Daily Mirror to the Yorkshire Post, Kent Messenger, and the Monmouthshire Beacon.
2. Between the Cairncross Review, the Furman Review, the CMA market study and recent mobile ecosystems study, as well as direct publisher experience, there is clear evidence that digital businesses – including national and regional news media – suffer from weak negotiating positions relative to platforms and aggregators and do not receive appropriate compensation for the production of reliable news content.
3. Among other applications, the NMA is confident that platforms and aggregators, including Google, are rapidly developing AI applications with the capability to source and utilise data sources on a mass scale. International text aggregators/database businesses such as ProQuest, LexisNexis and Factiva are positioning themselves as the data repositories to service this market.
4. News media content is a key input to commercial text and data mining (“TDM”). To quote the promotional material of one TDM business, Lexis:

*“Data lies at the heart of digital transformation. Whether forecasting with predictive analytics or conducting trend analysis, improving competitive intelligence or ensuring accuracy and compliance.....Use global, current and historical news content, company and financial information, patent data and more to identify investment opportunities and emerging market trends.....Stay up to date and better protect your brand with insights from global news coverage and social commentary and real-time alerts.”<sup>1</sup>*

5. We are concerned that government proposals to strengthen TDM exceptions could undermine commercially licensed relationships that exist between news media and the research & information sector and at the same time embolden the unlevied use of news media data by a whole range of other businesses, including global platforms, corporate insight, trend modelling, data processing and risk management, that are seeking to utilise such data without paying suitable compensation.
6. The reality is that while news publishers have expressed strong concerns about the ‘take it or leave it’ terms imposed by dominant search engines such as Google, in relation to text and data aggregation, the government’s proposals risk enabling a much broader swathe of companies to plunder news media data without paying equitable compensation. The NMA is

---

<sup>1</sup> <https://www.lexisnexis.com/en-us/professional/data-as-a-service/daas.page>

aware of an ever-widening range of businesses that are rapidly developing AI applications with the capability to source and combine news data sources on a mass scale.

7. Among them, for example, brand safety vendors such as Integral Ad Science and Double Verify (which crawl our members' sites to prevent ad misplacement) are using their position as a way to further crawl news publisher sites in order to target advertising based on keywords or emotional sentiment. In doing so, these vendors are building audience segments that they are able to sell to advertisers, without a clear legal right to do so. The effect of this mass use of news publisher data is to undermine the ability of news publishers to build and sell advertising inventory on their own terms.
8. As the IPO will be aware, many news publishers operate websites that do not require sign-in to a paywall, enabling open access according to a set of terms and conditions. Crawling of news publisher websites for commercial purposes, in the absence of a commercial licence agreement being in place, is extremely difficult for news publishers to prevent. The 'user agents' that carry out such crawling are hard to disambiguate from individual users, or the user agents that are required to carry out legitimate tasks, such as the brand safety role that is required as a result of contracts with advertising agencies. Thus news publishers' only route to prevent such commercial crawling is through the enforcement of commercial terms, with a view to establishing a legal contract. A broadened TDM right would embolden vendors to crawl news media sites without a contract, the result of which would be to undermine any ability for news media publishers to take action, further imperilling their existing and future revenue streams.
9. More directly, news publishers have earned revenue from aggregators by negotiating a royalty against prorated income based on search results and end-user subscriptions but this model does not translate easily to TDM where fees are levied for bulk access to content APIs. The Copyright Clearance Center, which is the US copyright collection agency for publishers, has built proprietary technology to support TDM and it forecasts rapid revenue growth. To date, they are licensing content from 60 publishers, mainly in the medical and science sectors, however they foresee expanding demand for news content. As rights holders, the news industry is only at the start of understanding the commercial value of TDM, and the increasing sophistication of derivative content and applications. We are concerned, therefore, that any broadening of TDM rights in the UK could undermine both existing and future revenue streams that are underpinned by TDM.
10. In circumstances where an AI algorithm or process learns its functions by ingesting large volumes of copyrighted material this must be adequately addressed by the law. As the American News Media Alliance has observed:

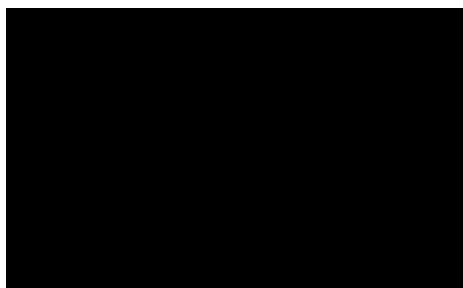
*"The content appropriated by these technologies, typically in massive quantities, may be used to train the AI to perform a variety of functions, which increasingly includes learning from news media reports how to write a story, and drawing on content from multiple sources to create a rendition of the news that is not identical to that of any one contributing source while being completely dependent on all of those sources in combination".*

These news products disintermediate between the source and its audience, encouraging the audience to remain inside the tech platform's ecosystem rather than link to the original information provider. AI 'trainers' must compensate the rights holders whose copyrighted works their machines are ingesting, as part and parcel of doing business. Licensing solutions are available which make works accessible at low cost on terms that respect the interests of rightsholders. We would be greatly concerned by the unlicensed use of our members' news

content by technology companies as raw material for machine learning (ML) purposes, since this could threaten the viability of the news industry.

11. The NMA advocates the preservation of the UK's narrow copyright exception for TDM that enables computational analysis solely for the purpose of non-commercial research. The UK should not be tempted to adopt a more permissive regime since this would unfairly skew the balance between the interests of IP owners and data holders on the one hand and those of data users on the other. In addition, there is a need to ensure that the enforcement regime enables rights holders to act against large-scale infringement, where AI developers have used copyright-protected works without authorisation.

We would be very happy to facilitate meetings with media organisations and their publishers, editors and legal advisers in order to give more detail on how they licence works for TDM and why fair payment for original content is vital to maintaining a healthy and diverse digital ecosystem.



7 January 2022