**25 August 2023**

**NUJ submission to the Communications and Digital Committee call for evidence on Large language models.**

The National Union of Journalists (NUJ) is the voice for journalism and journalists in the UK and Ireland. It was founded in 1907 and has more than 30,000 members working in broadcasting, newspapers, news agencies, magazines, book publishing, public relations, photography, videography and digital media. The NUJ is submitting evidence as its members’ content is used to inform large language models, with their development impacting journalists and journalism. The union is not affiliated to any political party.

**Capabilities and trends**

**1. How will large language models develop over the next three years?**

Large language models are likely to develop at pace in coming years, as AI developers recognise the profitable opportunities systems bring. Government’s AI white paper notes “In the last 4 Years, LLMs have been developed beyond expectations.” As Ministers and employers seek to understand various ways LLMs can be used, tech companies will likely develop systems further, refining their existing ability to produce content using data input, improving efficiency. This will exacerbate existing concerns over threats to journalism and across creative sectors, particularly where data used to inform systems is inaccurate or accessed without consent.

The NUJ would welcome engagement with Ministers on the benefits and risks posed by LLMs and AI more widely.

**a) Given the inherent uncertainty of forecasts in this area, what can be done to improve understanding of and confidence in future trajectories?**

Although government’s monitoring and evaluation of the overall regulatory framework and how principles are applied is welcome, engagement with the NUJ and representative bodies should be included as part of this process. The Impact of LLMs on journalists and journalism cannot properly be understood without discussions with relevant stakeholders.

Horizon scanning should be well- funded and resourced to remain abreast of developments, and government must consider a statutory approach to regulation if monitoring identifies poor adherence to guidance, gaps in regulatory oversight or trends that indicate a need for legislation.

**Transparency**

**2. What are the greatest opportunities and risks over the next three years?**

When used ethically and with transparency, AI presents opportunities for good. As LLMs use large data text sets, within journalism there is grave concern over the scraping of content without due regard for the moral rights and copyright of individuals. Informing systems in this way poses harm to journalism and wider society, even if it is unintended.

Risks include AI generated text using existing journalistic content without the consent of copyright owners or text producing content based on inaccurate information input into systems. Journalism underpins democracy, relying on public trust and confidence in information accessed. Without clear labels identifying text generated by AI as such, there is an increased risk of both misinformation and disinformation across society. LLM developers should hold licences for information used. In addition, developers must be required to disclose this information within agreed timeframes.

Without new, accessible routes to redress for rights owners where breaches occur, there is a risk AI developers will continue to benefit economically, while those whose content is improperly used face arduous legal battles they cannot afford. The imbalance between developers and freelances for example, is great - one way to level the playing field is by ensuring there are transparent processes in place that developers are compelled to comply with.

Government’s AI white paper recognises that public trust in AI is critical as otherwise there will be a reluctance to use it. The NUJ emphasises that developers have a crucial role to play in maintaining public trust by abiding to ethical principles including those impacting equalities concerns including bias and diversity, as well as those outlined in government’s paper. There must be records kept of original content used to inform systems, to ensure developers are held to account.

Government’s approach relies on regulators producing guidance and applying principles within their remits. As the principles will be issued on a non-statutory basis, regulators will use their discretion on their application, assessing their relevance. Yet the fast-paced nature of AI and development of LLMs require clear frameworks that hold developers to high ethical standards with clear sanctions for breaches. There is a risk that government’s approach focuses on innovation at the expense of ensuring ethical AI practices are in place. This is to the detriment of journalists and other rights owners.

While the introduction of a duty to have due regard would require regulators to demonstrate they had considered the principles, it is unclear what sanctions would follow if breaches occurred. The profiting of big tech at the expense of journalists including freelances without financial backing to challenge rights infringements is not suitably addressed through government’s approach.

Engagement with the NUJ is welcome, to uncover opportunities within journalism that do not seek to replace skill sets held by journalists, but instead use systems in a manner that complements capabilities. There is great danger in having LLMs recreate and produce content without the style and diversity of thought held by humans. Although systems may be able to replicate the likeness and style of journalists and creators, the individuality that comes with journalism by humans cannot be replicated in the same manner.

**a) How should we think about risk in this context?**

Both short and long-term impacts of risks outlined above must be considered. Although LLMs may support journalism when used ethically and lawfully, there is a long-term risk of harm to the profession if there is a loss of talent arising from job cuts, as publishers seek to use systems in cost-cutting efforts. The experience and skill held by journalists cannot simply be replaced by algorithms reproducing at speed. Some publishers are adopting the use of AI without engagement with staff or unions to diversify ways of working that increase profits, without consideration for ethical concerns.

**Domestic regulation**

**3. How adequately does the AI White Paper (alongside other Government policy) deal with large language models? Is a tailored regulatory approach needed?**

A tailored regulatory approach that considers the far-reaching impact of large language models is required. The AI White Paper rightly considers examples of risks including to safety for example, if advice offered is based on inaccurate information scraped from the internet. This illustrates the importance of requiring content generated by AI to be labelled as such – improving understanding and instilling confidence in those accessing material.

The NUJ believes a balanced approach considering the benefits of AI without undermining the rights of copyright owners is the most appropriate action for government to adopt. We note industry feedback referenced in the White Paper about the harm a rigid approach would cause, but caution against the non-statutory approach placing pressure on regulators who must exercise their discretion on what guidance they deem necessary to issue. We welcome an opportunity to put forward views on behalf of journalists, whose material will be used to inform LLMs.

**a) What are the implications of open-source models proliferating?**

Due to the structure of open-source models, there is concern that their proliferation will make it increasingly difficult to identify those responsible for rights breaches – for example scraping content without consent to inform systems. As structures are more complex than other models, there is a risk data used to train systems can become biased if large amounts of harmful, discriminatory content is used throughout chains.

**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?**

Regulators have themselves expressed concern about lacking the statutory basis to apply principles. There is no indication beyond “when parliamentary time allows” of when a duty requiring regulators to have due regard of the principles will be enforced. Again, this risks sending a weakened message to developers that strengthened legal frameworks are not a priority for the UK government on AI.

Government states the “balancing of the rights of content producers and AI developers” is outside the scope of proposals for a new overarching framework for AI regulation, but the NUJ considers this an essential consideration of improving the effectiveness of any regulatory approach.

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

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

Engagement with the NUJ and representative organisations of individuals whose content will be used is a necessary non-regulatory approach easily adopted by government. There must be a willingness to hear from a range of stakeholders to gain an accurate understanding of concerns and opportunities. Any efforts by developers to exclude trade unions/representative bodies from engagement with government must be met with opposition, instead recognising the benefits of hearing journalists’ perspectives.

Government must consider legislation to underpin principles, avoiding a voluntary non-enforceable approach of guidance. In addition, regulators must be well-equipped to respond to developments and government’s central functions body must be reactive where gaps in regulatory oversight are uncovered. Innovation can be achieved alongside compliance, making the UK a good example within the AI landscape.

**b) At what stage of the AI life cycle will interventions be most effective?**

It is crucial interventions occur as early as possible to increase their likelihood of success. Developers must be aware of the legal and ethical standards they are expected to adhere to to train and develop LLMs. As some AI risks will not be covered by existing legislation, LLM developers must be clear on consequences of poor practices – for example in perpetuating discrimination and biases. As developments on AI are frequent, approaches to interventions must be both proactive and reactive, adopting flexibility to address new concerns as they arise.

**c) How can the risk of unintended consequences be addressed?**

The NUJ has called for new, accessible, and clear routes to redress when content is used without consent. Fair compensation accessed through structures would allow journalists to receive fair remuneration. It is disappointing therefore, that government’s approach will not include new routes to redress.

**International context**

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

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

**b) What is the likelihood of regulatory divergence? What would be its consequences?**

Any approach to LLMs regulation should consider the UK’s international influence and the risk of other countries adopting weakened positions if action by Ministers fails to adequately protect journalists’ rights. Divergence risks making the UK an attractive place for developers advancing systems without the ‘constraints’ of statutory frameworks with sanctions for breaches, and accessible routes to remedy infringements where they occur.

**National Union of Journalists**