

Dr Ingo Frommholz, Reader in Data Science, School of Engineering, Computing and Mathematical Sciences, University of Wolverhampton; Sebastijan Maček, Journalist/Translator, Slovenian Press Agency; and Dr Chris Wyatt, Research Impact & Policy Manager, University of Wolverhampton—written evidence (FON0031)

**House of Lords Communications and Digital Select Committee inquiry:
The future of news: impartiality, trust, and technology**

Dr Frommholz is the head of the Data Science and AI group and an active researcher in Information Retrieval, AI and Natural Language Processing (NLP).

1. New Technologies have provided news agencies and broadcasters with unparalleled real time news and analysis. With these technologies have come methodologies that, while promising speed, do so according to algorithms that provide issues of their own. This submission will show how they contribute to declining trust with concomitant concerns regarding impartiality and offer some suggestions to improve the position.

Questions on Trends

Qu1

2. We see these as operating on three levels: algorithms and technology; language; and culture. In the first of these, the way an algorithm is written will yield different results. To take a simple example, average salary could be assessed by mean, mode, and median, all yielding different results. Scaled up to the news environment, algorithms can dictate the order in which news items are seen or even omit news items altogether. Key characteristics of a news story could be removed or distorted by the way information is portrayed. The main way in which large platforms and aggregators affect the news environment is by typically rewarding opinionated, biased content in favour of the kind of impartial and in-depth reporting that had been characteristic of large legacy news organisations in the past. Algorithms have forced all news providers to produce sensationalist stories. Such stories tend to get more traction and the way purveyors of quality news have responded is by becoming more tabloid-like. This explains the prevalence of clickbait, which has undermined quality journalism.

3. Technology has also lowered entry barriers; setting up a web page is sufficient to call oneself a news outlet. This is not necessarily bad since it has given rise to important new formats, like the open-source intelligence Bellingcat but the flipside is that outlets that used to bring a wide variety of topics now expose readers to a narrower selection of topics. Focussing on a narrower selection on topics (sports, culture, society) might mean the reader misses out on other important issues, such as politics. Often smaller outlets pick up a story and package it for their audiences or put their own unique angle on it. More intensive work is still being done by legacy media such as the BBC, whereas

these smaller outlets piggyback on their work. In the long term, this is not a sustainable model and once it collapses, which it will sooner or later, it will be hugely detrimental to democracy. Most importantly, technology platforms have severed the link between advertiser and media. Media used to be the main way in which advertisers interacted with their customers. Platforms have destroyed that, advertisers no longer need media to reach audiences, they can go straight to the platforms. Historically successful revenue models (Guardian, FT, Economist) no longer function as they did. In this sense, journalism is not just any business and the need to nurture debate and democracy cannot simply be left to the market.

4. Language, too, has an effect in this process. Large Language Models (LLMs) and generative AI are discussed in more detail below but have a bearing here. The emphasis for the news is the way the language may be used to obfuscate the actual story, turning the emphasis from one aspect, eg crime, to another, eg immigration. This aspect can be key to the functionality of an algorithm and there may be legal implications to what may be said and how. We have been involved in works utilising LLMs for instance for authorship attribution and authorship obfuscation, the latter being important if sources have to stay anonymous.

5. The third aspect is culture and we have been involved in studying how Hofstede's six cultural dimensions can be and have been applied to website design. These are: the distance from power; the interface between individualism and collectivism; masculinity and femininity; the degree of avoidance of uncertainty; long- and short-term thinking; and the imperatives of indulgence and restraint. These aspects are common to all cultures and enable sensible cross-cultural analysis to take place. We have found that cross-cultural dimensions can be used to model users' preferences on search interface design and are conducting further research vis-à-vis content. Application of this research on news outlets is important future work.

6. All three of the factors above underpin news plurality. This is unlikely to change and trends may, indeed, become more accentuated from where they are now.

Qu2

7. There are two main aspects here: the generation of potentially fake content, and its proliferation for instance with search engines.

8. At its current stage, generative AI will not automatically produce content that renders journalists obsolete. AI cannot in fact produce new content as such, it needs input, be it human or automatic, and its output is typically below standard and requires further editing, though this is bound to improve as technology evolves.

9. The deeper question here is the notion of what is real. The recent Taylor Swift deep fake porn is a good example. With AI producing such high-quality images, video and audio, there is no way for the average consumer to know whether something is real or not anymore, even at this early stage in generative AI development. The problem is bound to get worse. Imagine a major war

speech by a prime minister being created by enemies and widely distributed online: even when discovered to have been fake, the damage has already been done. People's trust may be eroded, opinions formed, fear instilled.

10. Modern AI is an order of magnitude more impressive from the early fake news of a decade ago. At present there is no effective way to deal with this, although there are ongoing important research initiatives in the information retrieval, AI and NLP communities to provide technological means to support fact checking and fake news detection. The AI, retrieval, and language technologies we are developing can readily be used for such purposes and will play a central impactful role in the future. Beyond the technology, the current discussion in the media is that respected outlets should become guarantors of reality of sorts but, given the declining trust in the media overall, this is not an entirely workable proposition, given the need for first-hand experience as the only guarantor of reality. If as a society we do not operate in an environment with even just a basic set of shared facts, how can society operate at all? A solution therefore has a non-technical dimension, it requires media literacy, and it should be supported by technology that supports transparent fact checking and fake news detection, giving the user more tools at hand.

11. Search engines and recommender systems, which utilise generative AI as well, are still the main technology in the dissemination of information. Our group has more than two decades experience in search engine technology and information retrieval. But search engines are also problematic for business models. This is because screening and refining options are increasingly taken out of the hands of both a given media business and of individuals seeking to access that information. This is particularly pronounced in search algorithms, where fake or generative AI content appears in a search ahead of good journalism (see the discussion on clickbait above). As things stand, this is having an injurious effect on media business models. This requires search and recommendation algorithms to integrate more quality signals into their algorithm that also consider the potential trustworthiness of information. Research on fact checking and fake news detection can be utilised here. Also, recent developments in AI and LLMs (applied, e.g., in the scholarly domain to determine qualitative output) can be promising for this purpose.

Qu3

12. Perceptions of impartiality are changing as society fragments along the lines, say, of intersectionality and identity politics, the effects of the post-pandemic economic situation, political parties, and differing perceptions of fairness. News organisations have to navigate these differences and are often accused of a lack of impartiality because they cleave to a particular position. The plethora of contending visions for the country's future and remedies for the issues of today cannot be put back in the box. The consequence is that positions taken will always be attacked by someone for being biased, slanted, or incorrect.

13. True impartiality is almost certainly impossible to achieve. Recent events in Israel and Palestine are evidence of this. Balance between contending viewpoints, presenting both, would be less problematic than trying to tread an impossibly narrow path. Throughout this submission, impartiality is being treated as actively different from balance.

14. Perhaps the term 'impartiality' should be abandoned in favour of 'closest possible approximation to the truth'. All media operate in an asymmetrical information environment, as they always know at best a portion of all the relevant facts. They will necessarily be incomplete, which means they cannot be impartial or unbiased.

Qu4

15. LLMs, such as ChatGPT, operate across various media and other domains. Our research demonstrates and highlights that there is a real concern regarding the effect of machine-generated texts. These include computer-generated texts that resemble human text in subject and written style, but also multimedia documents such as video and images (see above). This has the potential to be extremely serious and the ability to be able to differentiate human-vs-AI-written text in all contexts is essential. Recent news stories showing how people are unaware that they are communicating with AI are illustrative of the problem. Deep fake imagery is a similar example of this trend.

16. Add to this questions of omission, cognitive bias, and cultural differences, there is scope for members of the public to doubt everything they see and hear. If this evolves unchecked, members of the public will look to other sources of information, undermining the state and its institutions. These could come from malign domestic actors, demagogues seeking power, or hostile states seeking to erode trust and legitimacy within.

Qu4a

17. The desired state would be trust in impartiality but the issue is increasingly becoming one where trust is being linked to partiality instead. Study after study has shown that people trust sources based on like-mindedness, kinship, tribalism, they trust sources that play to their preferences, opinions and biases. This is because the worldviews informing viewpoints are too far removed one from another. This is very difficult to reverse, especially at the moment where middle ground is in short supply.

Qu4b

18. Information is trusted or not because of trust in the expertise and rigour of the person or organisation imparting the information. Where trust is lacking information, regardless of its veracity, it is considered to be dis- or mis-information. Analogously, this happens with academic peer review and a comparison with that process would prove instructive. As we put it in a conference paper last year:

'With millions of research articles published yearly, the peer review process is in danger of collapsing ... Challenges arise from the large number of manuscripts submitted, skyrocketing use of preprint archives and institutional repositories, problems regarding the identification and availability of experts, conflicts of interest, and bias in reviewing. Such issues can affect the integrity of the reviewing process as well as the timeliness, quality, credibility, and reproducibility of research articles.'

The same is the case with the news environment: millions of stories produced not entirely dissimilarly, with similar conflicts of interest and questions of bias. The net effect is the same on 'the timeliness, quality, credibility, and reproducibility' of news stories and features. The point is that the broad process is similar across many spheres and the issue of diminishing trust is, likewise, common to them all.

Questions on Evaluation

Qu1

19. News organisations have been adapting their business models for the last decades, but they are unable to respond effectively because of the speed of the technology. The main changes needed to get out in front of the pace of accelerating change are a) to manage the broad process of critical engagement; and b) to make changes to the programs/algorithms being used. If these two things are not done, the sector will find itself replaced by a mechanism of pre-selected one-line news on mobile phones.

20. Outlets such as BuzzFeed or Vice in the US that adapted amazingly well to social media distribution were thrown into disarray or bankruptcy once social media started deprioritising news in their feeds, depriving them of traffic. In the creative destruction of the field, there are now very successful newsletter-based outlets and podcasts that have adapted amazingly well and are serving their audiences with high-quality journalism.

21. The downside of this is that deep organisational knowledge is being lost. Much like a university or any other organisation, a news outlet develops internal know-how and deep historical knowledge. This is why outlets such as the Economist can draw on decades of knowledge and provide very nuanced, complex reporting on major stories, reaching back to find parallels in the past, and learning from history. New media have only a few years before the next technological development throws them into a disarray and have no way of forming such deep knowledge, which is crucial to the functioning of serious media and democracy.

Qu2a

22. We would argue that they cannot, as the question is phrased, because 'balance' is being synonymised with 'impartiality'. A clear example might be the dispute over Kashmir. Impartiality is not taking a side. Balance is an appreciation and understanding of the contending arguments put forward by the two countries involved. We argue that balance is a better mechanism for trust than attempting true impartiality. Balance, too, is a better mechanism for setting out values and defending them. Setting out values and then arguing that this is somehow impartial is neither logical nor balanced.

Qu3

23. There are two, interlinked, questions here. The first is to tackle disinformation and the second is how this functions when a section of society

reacts against it. The former may relate to the veracity of the information, on the one hand, and how it is understood, on the other. News media organisations try to ensure that they do not communicate untrue information. Information that is complex and open to interpretation is another question. Information and understanding relating to, for example, the foreign policy of Iran is nuanced, especially with proxies. Veracity of information on decisions taken behind closed doors is difficult to achieve, with ascertaining associated motivations harder still. It should also be mentioned that there are several sources of disinformation, for example social media. If, for instance, a random post by a content farm on Facebook starts sending out disinformation, there is the question of whether the media should really be the ones to dedicate precious resources to debunking it. This is only feasible if it is relevant to their core reporting.

24. With regard to the second aspect, tackling mis- and disinformation will inevitably alienate some section or another of society. As above, a position of balance is necessary, as any attempt to be impartial will not work because balance enables refutation in a way that impartiality cannot.

25. It should be noted that the definition of disinformation is sometimes straightforward, but often is not. It is a label that is even more abused than impartiality and is often used by people in positions of power to silence dissent and silence the media by accusing them of peddling disinformation.

Qu5

26. The Government should enact legislation to regulate the governance of AI. It is not workable to micro-manage the development of the software but it is a reasonable proposition to regulate its functionality and hold those responsible for it to account. Good examples are LLMs but also large multimodal models (LMMs), which are capable of creating text, images and videos from textual prompts. The quality of their output will improve and develop over time but the key issue is to differentiate them from human-created content. This latter consideration is the key aspect upon which legislation would be required. Legislation should be accompanied by research on technological solutions, supported by funding bodies such as UKRI and others. By analogy, for example, Germany has provided tremendous funding supporting research on tackling disinformation.

Qu5b

27. Three aspects are essential for the Government to grasp and for action to be taken: the first is that the relentless pace of mis- and dis-information should be matched by the Government. It often appears that the scale of mis- and dis-information, too, is not matched. Too often, Government reacts to a hundred pieces of mis- and dis-information with a single press release and then thinks that the job of refutation has been achieved, when it clearly has not been.

28. The second is to ensure that information, when it does appear, is the whole truth, rather than showing bias by omission. For information to be trusted, ways need to be found to prevent a narrative that twists information around. This problem is made more difficult when the original narrative misses out

information, thereby creating and fostering a narrative of either duplicity and mendacity or complacency and incompetence. Overcoming this problem enables a more balanced discourse.

29. The third is that there needs to be education and understanding that the information being conveyed may be, variously, incomplete, slanted, and biased and how to look for multiple sources of information and develop understanding. As the Call for Evidence makes clear, 'the UK faces a general election amid fears about AI-enabled mis- and dis-information'. In an atmosphere of recent advances in large language models, the impacts on media plurality, and the use of generative AI tools becoming more widespread, education and political, social, and economic literacy will go a long way to preserve an informed electoral process, both now and in the future.

12 February 2024