AI in the newsroom: implications for educators from an experiment with trainee journalists

By Sean Tunney, Adam Cox, Athanasia Batziou and Yuwei Lin, University of Roehampton and Imperial College, London

Abstract

The adoption of Artificial Intelligence (AI) technology in the newsroom has raised pointed questions about its effects on journalism quality, professional practice and industry incentives. Risks include legal and ethical challenges, as well as the potential for AI to introduce inaccuracies, bias or misinformation. Less attention, however, has been paid to how journalism educators will need to respond. Using an experiment designed to capture the thinking of so-called digital natives when working with an AI tool, we identify emerging issues educators will need to address from human-machine collaboration. We suggest three areas of concern: the extent to which younger journalists will be sceptical about AI outputs, the agency journalists using AI will have, and risks associated with the lack of transparency inherent in AI tools. We offer initial recommendations for educators.

Introduction

Since the emergence of auto-writing tools for journalism in the 2000s (Chopra, 2022), researchers have explored questions about bias, information quality, industry dynamics and various issues related to artificial intelligence (AI) in the newsroom.

Given advances in natural language generation (NLG) and related technologies, the critical questions have become more acute (c.f. Sundar & Liao, 2023). It's a simple proposition: As the tools become more powerful, the potential for radical or far-reaching effects on the industry and society increases. Researchers have expressed concerns that legacy media companies could end up ceding control over information selection and access to some news sources, eroding their watchdog function, as dependence on generative AI expands (Simon, 2022).

Following the November 2022 release of OpenAI's ChatGPT, discussion about the impact of auto-writing tools has spilled out to the wider public. Suddenly, a powerful, useful and, at times, amusing text-generating tool has become available to millions of people around the world. The notion that human journalists could be replaced by code – if not entirely, then at least in large numbers – looks more plausible than before.

Commercial incentives for news providers may reinforce such concerns. Machine-generated content is cheaper and faster to produce than human journalism (Miroshnichenko, 2018), and numerous media groups have deployed or experimented with text-generating tools (Simon, 2022). Initially, the AI was primitive, using templates and data feeds to produce sports reports, election results, financial news and weather updates; tools were later developed that could generate short reports or precis from longer texts (Danzon-Chambaud, 2021). In recent years, news organisations have experimented with more sophisticated technology to create complex texts without human intervention apart from the AI's initial training (Floridi & Chiriatti, 2020).

Yet, despite the abundance of research on AI and journalism aimed at understanding these developments, there has been less devoted to how we will prepare journalists for an AI-saturated workplace. It is clear from the literature that the technology presents both risks and opportunities for journalists and their employers. Researchers and journalists have noted the lack of specific higher education courses in some countries addressing these issues (Pinto-Martinho et al., 2022). Educators, according to Pavlik, 'should be considering how to develop courses or programs that train human students in the effective use of generative AI, as well as the threats it poses, including matters of ethics and potential bias' (2023, p.92).

Academic discussion about AI and journalism has often focused on human-machine collaboration, as opposed to the idea of machines replacing humans. Indeed, no one has figured out a way to send ChatGPT into the field to cover a disaster or conduct a live interview. We are also focused on collaboration rather than replacement. In fact, the expectation that we will even need journalism educators in the future rests on the premise that journalists will work with machines rather than be replaced by them.

What has been rare in the journalism education research related to AI produced to date, is experimental research that could provide evidence-based findings on the emerging risks that journalism educators need to consider. Just as the advent of the internet and social media has required a re-evaluation of how journalism law and ethics are taught, we expect AI will require new thinking for teaching many aspects of journalism practice. In this paper, we explore risks associated with AI-assisted journalism, looking specifically at the possibility of a new generation embracing AI in ways that potentially create legal and ethical issues, or which could compromise journalistic standards and norms.

We first review literature on the legal and ethical risks journalism faces when utilising AI. This provides context for two research questions concerning the metacognition of trainee journalists and the potential legal and ethical risks arising from the use of AI in a practical setting. We then describe an experiment we developed to identify those risks based on the reactions of aspiring journalists, in which they are tasked with writing two articles, one with an AI tool and one without. In section three, we present the results and analysis of that experiment. Finally, we discuss the implications of AI-related risks and make recommendations for how journalism educators can address them.

Ethical and legal risks

AI tools raise pointed ethical and legal threats in journalistic production as the outputs can be inaccurate and biased. AI systems are often described as 'black boxes', with little information provided on data sources or decision-making. It is notoriously difficult to understand the complexity of algorithms and the data used to train some advanced tools (Bathaee, 2017). Among the issues for journalism and society, beyond transparency, are: bias and objectivity; accuracy; privacy (in both an ethical and legal sense); libel; copyright; and contempt of court (Kothari & Hickerson, 2020). There are also power relations to consider when users interact with AI. The moral agency and relative autonomy of individual journalists may be diminished whenever media organisations and platform providers govern the tools' usage (Dörr & Hollnbuchner, 2017; Simon, 2022).

Ethical risks: Bias, objectivity, accuracy and transparency

In terms of bias and objectivity, there is a divide. Researchers have shown some news producers themselves believe AI can reduce bias, while others have questioned this assumption (Montal & Reich, 2017; Thurman et al., 2017). Though we are focused on production rather than consumption, it is worth noting that news au-

diences may trust algorithms' objectivity, accepting 'practical and symbolic assurances that their evaluations are fair and accurate, free from subjectivity, error, or attempted influence' (Gillespie 2014, p.179, quoted in Montal and Reich, 2017, p.830). However, researchers suggest algorithms can change according to business decisions by corporate providers (Montal & Reich, 2017). These could embed unintended biases — where the unconscious preconceptions of a predominately white, male and elite workforce become immersed within the code, or where bottom-line factors eclipse concerns for fairness or equality (Black, 2023). Investigators have already highlighted how gender and race inequalities have been replicated by algorithms (Kothari & Hickerson, 2020; Mayson, 2019; Noble, 2018). Biases also might be intentional, based on marketing or other interests. Montal & Reich (2018, p.830) envision the 'algorithmic potential to affect the visibility of political and social actors while maintaining a predetermined agenda'.

Fused with this in the literature is the question of AI introducing inaccuracies, blurring fact and fiction, or conversely helping to make news more accurate. AI has been found to help in fact-checking and verification. Some journalists told researchers they believed, mostly, that, as a software manager put it, the 'data doesn't lie' (Montal & Reich, 2017). Analysts, nonetheless, have suggested AI has been less effective in stifling misinformation. Instead, it may be more capable of disseminating disinformation than stemming its flow (Aïmeur et al., 2023). Where poor coding or data sources are propagated: 'Bad data can lead to bad results and automation can amplify the speed and scale of their dissemination' (Kothari & Hickerson, 2020, p.213).

A key issue for us is transparency. It speaks to questions about journalistic practice and the imperative to understand where information comes from. Transparency can be an issue on its own, or a factor in other ethical and legal challenges. For instance, not being able to understand how an AI system produced a text can create or exacerbate the risk of libel, privacy, contempt of court or copyright offences. Many AI developers may have reason to keep their algorithms hidden, but this acts against clarity in the news process. Transparency acts as an ethical principle for journalists, and publishers often promote it throughout the newsgathering process. For AI-assisted journalism, this issue is complicated by the degree of algorithmic transparency — the methodology, construction and limitations of an algorithm. Potentially opaque issues include information quality; benchmarks for accuracy; uncertainty; timeliness; completeness; data collection assumptions; sources and training data; alongside human influence (Diakopoulos & Koliska, 2017).

Legal risks: Libel and privacy

Turning to legal considerations, news organisations that source AI will not be able to rely on the same libel defences tech organisations have traditionally utilised when deploying algorithms to produce content. A cruel truth that journalism students are taught is that every defamatory statement reporters may unwittingly publish is a fresh libel (Hanna & Dodd, 2020). The risk is exacerbated by pressure on news providers to cut costs and speed up production, as advertising revenue has fallen, leading to more limited sub-editing (Juntunen, 2010). Libellous material sourced from AI, when not spotted before reaching the page, can be costly to bottom lines and reputations (Ombelet et al., 2016). A less clear and present danger arises if AI scrapes data that are prejudicial to a present or future trial; the publisher could be liable for contempt of court in some parts of the world (e.g. Hanna & Dodd, 2020; Kothari & Hickerson, 2020).

Another issue is privacy. AI does not 'know' whether people's personal data that it supplies were acquired for another purpose, i.e. advertising or via profiling. This could be an issue under General Data Protection Regulation (Ombelet et al., 2016). There also may be copyright issues in some jurisdictions. AI might source protected material that could be reprinted as original for commercial gain (Hanna & Dodd, 2020).

Looking to the future: 'Digital natives' and AI

Why centre the research on students, aside from having an educational focus? Scholars have interviewed journalists to determine their views on AI impacts, including the effect on employment, concerns such as transparency, accuracy and bias (Thurman et al., 2017) and professional values (Komatsu et al., 2020). They have sought to find out what skills are required by industry and argued that universities are uniquely placed to provide them (Bucknell, 2020). They have learnt from educators how journalism students might be taught to train in these skills and work with the new technology (Gómez-Diago, 2022; Kothari & Hickerson, 2020). This has suggested the importance of educating students in creating automated IT content, alongside media literacy, source validity and fact-checking methods (Gómez-Diago, 2022). Nevertheless, we were interested in a particular premise, based on the question of whether so-called digital natives (Prensky, 2001) might be imbued with particular skills, such that they can search and navigate digital media more effectively than older

counterparts. While a generalised version of this notion is widely influential in journalism education (Bethell, 2010; Matsiola et al., 2019), other investigators have found less support for this idea (Nygren & Guath, 2019, 2022; Wineburg & McGrew, 2017). In any event, there is evidence young people hold this belief, considering themselves as skilled fact-checkers when they are not (Nygren & Guath, 2019). Indeed, there is a debate as to whether youngsters who believe their aptitude to be greater experience an overconfidence akin to the 'Dunning-Kruger effect'. As David Dunning describes the phenomenon, 'their incomplete and misguided knowledge lead[s] them to make mistakes but those exact same deficits also prevent them from recognizing when they are making mistakes and other people [are] choosing more wisely' (Dunning, 2011, p.248; Nygren & Guath, 2019, 2022).

Researchers have identified particular blind-spots. Younger people in one study failed to differentiate advertising from editorial, or discern what experienced journalists might describe as advertorial, i.e. messages with commercial intent masked as impartial news or statements (Nygren & Guath, 2019). Other investigators discovered youngsters had difficulty unmasking covert ideological intentions. They compared the results of professional fact checkers with those of young people and found the scepticism and methodical techniques displayed by professionals proved better in discerning misinformation (Wineburg & McGrew, 2017). In terminology made famous by Donald Rumsfeld, we suggest fact checkers were more wary of the possibility of 'known unknowns' and even 'unknown unknowns' (Kirk, 2016, p.110).

There is a suggestion that all age groups use mental shortcuts, or heuristics, to make sense of AI information. In one study (Sundar & Liao, 2023), when readers were told a report was generated by an algorithm, this encouraged a positive machine heuristic. That is, they viewed it as typically more objective than human-created content, even when it contained arguments that countered their own beliefs. Researchers have found AI reporting could also trigger a negative machine heuristic, where readers view the tool as too rigid for nuanced judgements, particularly when producing content other than 'hard news'.

There is also research that suggests some users have considered computers as a psychologically salient source, treating their interactions as if they were with humans. The natural-language response of an AI tool means that it could be viewed as such a source, including by journalists – providing the raw material for stories. Yet, its learning process means it is assembling pre-existing content (albeit in unpredictable ways), rather than offering original material. It is, as notoriously described, a 'stochastic parrot' (Sundar & Liao, 2023, quoting Bender et al, 2021).

Research questions

When we considered 1) the array of legal and ethical challenges, 2) the question of attitudes among digital natives, and 3) the notion of machine heuristics, a broad question emerged regarding the metacognition of younger journalists when using an AI tool. That is going beyond considering students' assessment of the status of source information and knowledge claims towards identifying their self-knowledge about the operation of this reasoning (c.f. Lee, 2022). We felt that, by exploring their thinking, we could identify risks associated with AI in a more granular and evidence-based way than had been discussed to date. We arrived at two research questions:

RQ1 – What ways of thinking are displayed by trainee journalists when using an AI tool to produce journalistic work?

RQ2 – What kinds of risks can be seen from these thought processes in terms of legal and ethical hazards or challenges to journalistic standards and norms?

In the next section, we discuss how we sought to answer those questions.

Experiment design

Our experiment was divided into three parts and lasted approximately two hours for each participant. Two parts involved role play and a third was devoted to semi-structured interviews, aimed at producing qualitative findings. In part one, individual participants were brought to a mock newsroom and asked to write a breaking news article using traditional methods. In the second part, they were to use an AI tool for a Q&A-style 'explainer' feature, of the kind often produced for controversial or running stories. Both tasks were time-controlled. In part three, they were interviewed about their experiences and views.

We ran each session individually, for two reasons. We were concerned that a group environment might create a competitive atmosphere and discourage introspection during the interview phase. Also, as this was a complex experiment with numerous steps, we wanted to ensure we could provide support for each participant. After piloting the experiment with one person, we recruited 15 current and recent journalism students, ranging from 20 to 30 years old. All of them were either in the final year of their degree or had graduated within the past two years; 11 were female and four male, with a range of ethnic and socio-economic backgrounds. This purposive sampling was a strategic decision (Hansen & Machin, 2019), based on the earlier discussion outlined on digital natives. There is no claim here for demographic representativeness, so, for instance, we do not consider the number of participants sharing a view as particularly significant. Those invited had to have demonstrated competency by passing modules addressing ethics on journalism programmes; beyond that, there were no prerequisites for participation. Participants were told they would be treated anonymously and were given a stipend to compensate for the time and cost of travel.

We chose Generative Pre-trained Transformer 3 (GPT-3) as our AI tool due to its ability to produce text resembling that written by humans and its capacity for learning (Wei et al., 2022). GPT-3 is similar to ChatGPT, however the latter was rejected for our purposes as its availability was unpredictable.

Our pilot prompted us to focus on digital natives' metacognition. This underscored the importance of selecting a news topic that could involve under-contextualised, misleading or incorrect information. We chose fracking. The topic can be considered hard news (it concerns economic policy, energy, the environment, national regulation and local government) or soft news (it receives attention from celebrities, including those popular with digital natives) (McCarthy, 2016).

The GPT-3 database was not up-to-date, so we set the role play in the past (Hughes, 2023). A 2019 government announcement and several campaign group statements were used for the breaking news task. This part of the experiment performed two functions; it allowed the participants to familiarise themselves with the subject matter; and it gave them a means of later comparing the experience of writing with and without an AI tool. In part two, participants were told it was several months later and their employer was testing an AI tool. They were asked to write an explainer about fracking. They were instructed in how to use GPT-3 and their 'editor' (played by a researcher) supplied initial questions to ask it, including questions about some of the actors, such as trade unions, where some views were out-of-date. Participants were told they could ask follow-up or different questions, and that they could use the internet.

Part three involved semi-structured interviews focused on decision-making, expectations of the tool's impacts and any concerns participants had. We sought to elicit views about whether their position as digital natives affected decision-making or views about the technology (RQ1).

There has been debate as to whether interview texts 'truly' represent events, beliefs and actions. Here there is a heightened concern with power dynamics. There is the danger of interviewees being self-aggrandising when interrogated by former or current tutors; participants could want to appear knowledgeable or be defensive. What is more, the questions were based on complex ideas and posed moments after interviewees had been writing on deadline. To counter this, we stressed that the participants' performance would have no academic bearing and we sought to create a neutral, non-judgemental and relaxed mood (Minkin, 1997). We did not review the articles they produced as this was not relevant for our research questions. We attempted to solicit information that considered metacognition through indirect questioning, with each member of the research team posing questions from a list of pre-agreed queries.

The interviews were analysed using qualitative content analysis in order to both read and code them (Mayring, 2014). We used NVivo software to capture indications of metacognition concerning a range of issues including trust, content quality (from multiple perspectives), productivity, risks, source claims and implications for the practice of journalism (RQ1 and RQ2). Further qualitative analysis was conducted to illuminate the data. This involved identifying salient textual illustrations, analogous to the 'anchor examples' in Philipp Mayring's explanation of what he describes as 'narrow qualitative content analysis'. Mayring defines these as 'prototypical text passages' within texts. They are relevant extracts, identified so as to describe or explain, exemplify or help itemise the thematic categories (Mayring, 2014, pp.88–94, 95, 97). We have labelled each of the interviewees as P1 etc in order to identify their individual ideas and comments.

Results and analysis

Two major themes emerged from the interviews: thinking about source claims and consideration on human-machine collaboration (RQ1). We then identified three areas of risk in terms of journalistic behaviour (RQ2).

What participants thought

First, we assessed how participants reflected on their assessment of the information coming from the AI tool (RQ1). This evaluated what were their 'known knowns', and whether participants considered these needed confirmation through fact-checking. But there was also an assessment of whether they considered information that they were aware of not knowing and that needed to be deduced through checking facts, let alone the more elusive 'unknown unknowns' (Kirk, 2016, p.110). We assessed how much they contemplated trust in AI information. We did this by initially asking indirect questions concerning their journalistic procedure and thought processes, providing an opportunity for them to comment unbidden on elements of trust, but aiming not to directly prompt them or point them in one direction or another.

We entertained the possibility of an effect akin to that of Dunning-Kruger here and a generational variant (Nygren & Guath, 2019, 2022). In other words, we considered whether the less that individual participants contemplated fact-checking, the more confident they would be that they, and their generation, would be capable in understanding the limits and fallibility of the AI tool. This was not a representative sample and therefore what can be deduced regarding larger patterns generally is bounded collectively. Nevertheless, we considered that assessing this illuminated individual metacognition.

Some respondents – in particular P12 and P15 - suggested blind trust of the internet was indeed a problem but not so much for them; they felt that that applied more to the generation just behind them, who were more fully immersed in using the internet. Nevertheless, a number of respondents (P9, P13 and P15) expressed confidence in their ability to understand the operation of the AI. P9 and P13 also compared the AI's text with what they believed they knew about fracking, not considering discrepancies when it came to unions. They felt they did not need to confirm knowns and did not contemplate unknowns. P13 said of the AI source material: 'I could believe it because I already knew. I don't have to go and check every single thing it says necessarily if I already have the answer because I've done it before.'

As for reflections on the Dunning-Kruger notion and this generational variant, we found some of those displaying confidence as digital natives (P8, P10 and P13) were indeed among those who trusted the material implicitly, accepting the information without querying it. Participants reported picking 'out the key pieces of information' (P10) or 'just taking what is relevant' (P8).

We identified a divide, nevertheless, among those with this confidence in perceiving there to be generational differences. One confirmed that the tool was convincing and would be for those of any age (P15). Another was among the most emphatic at points that digital natives had special knowledge and linked the two thoughts. Thus, P10 said: '[A]s someone who has grown up with that sort of technology, I felt not only comfortable, but also safe in the fact that the information I was getting or the process that I was a part of. There are no issues from that perspective.'

However, others who exhibited confidence were more wary. P14 contemplated what we identified as a conception of known-unknowns and fact-checked concerns including about trade unions. Meanwhile, P2 considered what we saw as the possibility of known-unknowns, but did not have time to query the information. This participant considered limited fact-checking would suffice. 'I think if I Googled and everything it said was right, I would feel comfortable using it in the future.' Meanwhile, P15, who expressed confidence in understanding the AI, was also 'very wary' of the material, again identifying a lack of transparency – 'purely because it doesn't cite its sources'. This subject described a fact-checking process that queried whether the union material was reliable. Thus, to answer our first question, the interviews indicated, overall, that this experiment did not afford any simple illustration of the Dunning-Kruger effect (Nygren & Guath, 2022).

Second, we assessed how participants considered their relations with the AI to see how they navigated risk (RQ2) and we considered this in relation to machine heuristics. The interviews revealed a tension between concerns about AI negatively impacting on journalists and journalism, and enthusiasm for what they perceived as greater productivity and qualitative benefits.

Participants voiced concern about jobs being lost. 'I think my worry is definitely you're getting rid of journalists,' P15 said. 'It frightens me a bit how capable it is.' The interviewee envisioned the technology being able to generate work akin to what they had just written, without human involvement. At the same time, respondents could picture themselves using the technology. A number touted various benefits, despite perceiving shortcomings in terms of content quality (P5, P7, P8, P9, P13 and P14). Some welcomed the prospect of AI systems generating angles and ideas, or providing background or detail for articles (P1, P2, P5, P7, P8 and P10). They reported amazement at the speed of GPT-3. In fact, some of those who worried about job losses

generally and their prospects personally were among the most impressed by the productivity of the writing (P8, P9 and P14). It was clear that 'because of the speed, I cannot compete with it,' P14 concluded.

Interviewees also reported on what we have interpreted as the authoritativeness of the AI. Its language projected a convincing and coherent conviction. So, P13 told us: 'I was surprised by how academic and convincing some of the responses were.' P14 even joked: 'I felt like it was much smarter than I was.' We noted that some subjects shared a belief that the tool's logic-based nature and ability to tap into large databases meant it could provide balance, perspective and heightened objectivity. The productivity gains that participants perceived only underlined this point. 'It can give you ideas,' P11 said. Another, P13, added: 'That's one of the biggest advantages, it saves time.' P15, while maintaining some distrust, was partly won over during the experiment. 'I think I didn't feel as wary as I thought I would…. I felt like I was writing faster, and I was able to bring a lot more context in. And I don't think I would've felt as comfortable or qualified with writing an explainer about fracking without that type of tool.' P3, P11, P12 and P14 remained less convinced, reflecting the division researchers previously noted among more experienced journalists (Montal & Reich, 2017; Thurman et al., 2017). P3 and P11 cited a lack of transparency in the tool and that they could not be certain of the validity of the unnamed sources, paralleling concerns in the literature (i.e. Diakopoulos & Koliska, 2017). This had ethical and legal implications for a number (P5, P10, P11 and P12), as previous writers have suggested (Ombelet et al., 2016). There was the possibility that people could be libelled 'because I don't have reliable sources. I'm just putting names and blaming them based on an AI tool. If that was found to be incorrect and I'm putting that in, that could be a dangerous game.' (P11).

We perceived that some seemed to understand their relationship with the generative AI as one-way. It provided information which they used in the limited time they had (P5, P6, P8, P9, P13). Others appeared to ascribe more agency to themselves (P10, P11, P12, P14, P15). They conceived of their relationship as being more like an interview with a source. 'Some of the other questions that I asked were a follow-up of what the response gave me. I tried to investigate more from that perspective,' P10 explained. However, this was not necessarily an indicator of trust in the AI, with those engaged in a dialogue detailing different positions on this factor. Based on this, we considered agency an important factor in assessing the risks of AI-assisted journalism.

We assessed how participants understood their expectations of interacting with the AI by considering whether the exercise had consciously or unconsciously triggered machine heuristics (Sundar & Liao, 2023). One reported treating GPT-3 as a sentient provider. Yet, perhaps given the experiment's process, there was a consideration of not being 'fooled' by it. 'If I didn't know that it was artificial intelligence, I wouldn't have known that it wasn't a real person who was writing it,' P14 said. Our analysis of more interviews unearthed the degree of personification participants consciously or otherwise identified, borrowing from research on assistive technology (Purington et al., 2017). One gendered the AI (Abercrombie et al., 2021) as 'he' (P7), others consciously ascribed human qualities (P8 and P14), while more described how 'they' (P1, P3, P10 and P12) or 'the AI' (P12 and P13) or 'GPT-3' (P10 and P11) 'gave' or 'provided' information or answers to questions.

We considered the question of whether participants treated the tool as sentient was involved in any tendency to overlook or make allowances for AI limitations, in the same way that people will make allowances for other people's idiosyncrasies or unwelcome character traits. We found, for instance, some identifying a degree of personification expressed a wariness due to the style of the text, its occasional lack of relevance, or its vagueness (P8, 11 and 12). We found the readiness of participants to use AI for providing context, background, or factual detail did not appear to consider those limitations. None, for instance, explicitly addressed the possibility that an AI tool's selection of facts could be the result of hidden bias in the algorithms underpinning the system.

Rather, P5 and P10 extolled the logic of the tool and saw it complementing human journalism, with no apparent appreciation for how the 'balance' or 'context' had been generated. Limitations were most clearly seen when participants envisioned the tool working not in collaboration but on its own. That triggered negative machine heuristics in some (P1, P2, P11, P13, P14 and P15) (Sundar & Liao, 2023). Concerns about journalistic quality in terms of the writing were often limited to the robotic, predictable tone and a perceived lack of nuance in the texts (P1, P2, P11, P13 and P15). Thus, P2 concluded there would be 'a loss of something' if it attempted to construct features on its own. Another, P13, identified the something lost as 'a human touch'. This interviewee concluded that the 'AI will never understand a human being' or subjectivity's role, as it 'relies on logic, not emotions'. However, for some this triggered an explicitly positive machine heuristic. This rigidity was seen as beneficial, as it facilitated objective journalism, making the algorithm 'good' at delivering hard news output (P13 and P14).

We concluded that the possibility of users anthropomorphising an AI tool – when coupled with perceived benefits in terms of content, idea-generation and productivity – could heighten the legal and ethical risks from AI. Borrowing from a technique used in literature, which has been applied to other fields including journalism (Nünning, 2015), we likened the user experience to that of someone listening to an unreliable narrator; the power of such fiction is based on the reader's temporary belief in the facts presented, a belief formed without essential knowledge about the narrator's own limitations. In the case of AI and the notion of positive machine heuristics, this unreliable narrator concept is ironic given that the attractions of such a tool stem in part from the perception it may be more reliable than humans. The implications of these positive heuristics will be discussed in the next section, where we consider the risks for journalism practice.

Risks to journalism practice

Having identified some of the self-described thinking during the experiment, we considered the risks such cognition might engender in terms of journalism practice (RQ2). We suggest there are at least three significant and inter-related areas of risk in interacting with this unreliable narrator (RQ2). We present them here as hypothetical scenarios. It was tempting to base these scenarios on which views came up most frequently. As before, however, we were mindful of our sample not necessarily being representative, making any discussion about the volume of comments potentially misleading. Instead, we focused on the types of issues and concepts expressed by participants. Three risk scenarios stood out.

Risk: Young users of AI tools may devalue scepticism and deprioritise fact-checking

At least three cognitive factors contribute here. One is the possibility of overconfidence among users who may believe digital natives have a special aptitude for spotting inaccuracies, bias or misinformation. A second is the trust placed in authoritative language produced by auto-writing tools. A third is the possibility that users prove to be more focused on benefits in terms of speed and idea generation than on the possibility of machine-generated errors marring their output. These factors may work individually or in combination.

Risk: Users' agency may be reduced due to over-reliance on machine-generated text

This is akin to the phenomenon of atrophy in muscles not used. Journalists are taught to constantly ask questions – of their sources and of themselves – as they seek to make sense of sometimes conflicting facts and perspectives. The evidence of positive machine heuristics seen in participants' responses suggests a risk that young journalists could become over-reliant on AI-generated text, to the detriment of their own potential to synthesise and stress-test information from multiple sources. This was shown to be a particular risk for newswriting, and the context and background that is included in features.

Risk: The journalistic imperative of transparency may be compromised

As journalists use and come to rely on auto-writing tools, there is a risk that the lack of transparency in such systems comes to be seen as acceptable. Among other issues, there is a danger that any biases in the system remain hidden. Clearly, some participants had qualms about not knowing where information emerged from; but when some of those same users made checks and verified information, their confidence in the AI tool grew. Such feedback loops could lead users to normalise the idea of not knowing where information comes from.

Researchers have already begun to grapple with some of these risks. The Institute of Media and English at Birmingham City has laid out six principles for responsible journalistic use of ChatGPT in terms of diversity and inclusion (Birmingham City University, 2023). These included a call to recognise the importance of source material and a separate plea to be transparent where possible. Both principles were reflected in the thinking of some participants. But the six principles also call for users to be aware of built-in bias in the system and to build diversity into prompts, both issues that did not appear to occur to those who took part. Lastly, the principles called for a generally sceptical approach to AI-generated material, and it was striking how much credence appeared to be given to AI-generated content in our experiment.

Sundar and Liao (2023) have expressed concern about weaknesses in AI-based writing tools and the loss of human agency. Their unease was motivated in part by the prospect of positive machine heuristics, despite evidence from a self-experiment that showed the kind of false information that can be generated by ChatGPT. Their postulation, based on past research, was that different tasks would be more likely to trigger either a positive or negative machine heuristic, influencing how users trust AI in different circumstances. Our research indicated that, also, participants facing the same task could interpret aspects of the task differently,

with varying heuristics triggered.

The Birmingham researchers stressed the need for guidelines, a point made by participants in our experiment. Developing guidelines, however, is not straightforward. The speed of technological change, industry competition, and the possibility of unintended consequences from rules or norms all complicate such endeavours. Moreover, there are limits to journalist agency here (Dörr & Hollnbuchner, 2017; Simon, 2022), which extend to challenging bias. Reporters can vary their questions to the AI, but the opportunity to discern bias from a black box remains limited. Nevertheless, we see a role for journalism educators to play here too. Such a role begins with consideration of the thinking displayed by those who will use those tools and the behavioural risks that are posed as a result.

Implications for journalism educators

An immediate implication for educators from this research is similar to what tends to be advocated in response to just about every digital advance. That is, the basics still matter. For instance, students need to be taught a range of interview techniques. They should be encouraged to not simply succumb to the allure of public relations in any shape or guise. Educators need to insist on the importance of fact-checking. No journalism course worth its name ignores the importance of teaching media law. Clearly, the advent of auto-writing only reinforces these ideas.

But there are other possibilities that are suggested by our research that speak to the risks we have identified. One is that students should be taught how, at least at a basic level, generative AI works – AI literacy. Rather than focusing on IT skills, however, what is involved, as with previous technological shifts, is to transmit an understanding of the impact of the changing journalistic process (Hannaford, 2015), with an emphasis on understanding the issues of AI. This has relevance in terms of both transparency and agency. An understanding of some of the pitfalls of AI systems can teach the importance of developing sharper antennae that will detect when a lack of transparency can create legal or ethical problems. We are not suggesting here that journalism instructors develop skills in coding and computer science; but they should become conversant with the underlying concepts involved in NLG-based tools and how they generate outputs. In terms of agency, having a rudimentary understanding of how a system works can clarify for journalists the need to actively engage with, interrogate and evaluate it.

Another implication is that educators need to ensure that students understand AI in the context of industry dynamics. Our experiment suggests students need to be alerted to the attraction of AI for their future employers, operating in high-pressure environments.

Finally, we think educators can consider expanding the teaching of editing and link this with real-time interaction with AI, as so much of the participants' interplay with the AI here was referred to as editing the material produced by the tool. Educators could replicate workplace scenarios as we have done and use such activities to prompt guided discussion of issues thrown up by AI. Such teaching methods may well need to be interspersed with the delicate task of ethical and legal teaching to encourage resilience in students to know when it might be appropriate to stand up to editors demanding instant copy, when diverse and expanded interviewing is required.

Conclusion

To summarise, the experiment unearthed diverse thinking among digital natives, displaying divergent machine heuristics. From this, we identified three significant risks from young, less-experienced journalists being called on to use auto-writing technology or deciding on their own to use it. Those risks concern the value journalists attach to scepticism, the agency they feel they have or need, and the degree to which they are willing to accept the lack of transparency that, so far, has been notable in AI systems. To be clear, this is hardly an exhaustive list of risks.

From these risks, we have identified four important implications. One is that the basics in terms of journalism ethics and law training still matter. A second is that educators, if they wish to address these risks, will want to develop and impart a rudimentary understanding of how AI systems work. A third is that educators should consider AI implications in the context of commercial factors and the environment entry-level journalists will be entering. And a fourth is that students will benefit from practical instruction, with workshops that replicate or are similar to the experiment we carried out. Such exercises can encourage journalists of the

future to think carefully about the AI systems we assume publishers will expect them to use.

Artificial intelligence may usher in transformative changes in journalism practice. But the challenges to educators, while significant, need not be too daunting. They require a readiness to engage with new technology and an appreciation for the thought processes of those who will be using it.

Bibliography

Abercrombie, G, Curry, A.C., Pandya, M, & Rieser, V. (2021). Alexa, Google, Siri: What are Your Pronouns? (arXiv:2106.02578). arXiv. http://arxiv.org/abs/2106.02578.

Aïmeur, E, Amri, S, & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. Social Network Analysis and Mining, 13(1), 30.

Bathaee, Y. (2017). The artificial intelligence black box and the failure of intent and causation. Harv. JL & Tech., 31, 889.

Bethell, P. (2010). Journalism students' experience of mobile phone technology: Implications for journalism education. Asia Pacific Media Educator, 20.

Birmingham City University. (2023). Generative AI Diversity Guidelines. Birmingham City University. https://www.bcu.ac.uk/media/research/sir-lenny-henry-centre-for-media-diversity/blog/six-principles-for-responsible-journalistic-use-of-generative-ai-and-diversity-and-inclusion.

Black, J. (2023, May 18). Pronatalism and Silicon Valley's Right-Wing Turn. Tech Won't Save Us. https://techwontsave.us/episode/168_pronatalism_and_silicon_valleys_right_wing_turn_w_julia_black.html.

Bucknell, I. (2020). Going digital, not dying out. Journalism Education, 8(2), 13.

Chopra, D. (2022, November 6). Writing in the Era of Artificial Intelligence. Analytics Drift. https://analyticsdrift.com/writing-in-the-era-of-artificial-intelligence/

Danzon-Chambaud, S. (2021). A systematic review of automated journalism scholarship. Open Research Europe, 1, 4.

Diakopoulos, N, & Koliska, M. (2017). Algorithmic transparency in the news media. Digital Journalism, 5(7), 809–828.

Dörr, Konstantin Nicholas, & Hollnbuchner, Katharina. (2017). Ethical Challenges of Algorithmic Journalism. Digital Journalism, 5(4), 404–419.

Dunning, D. (2011). The Dunning–Kruger effect. Advances in Experimental Social Psychology, 44, 247–296.

Floridi, L, & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30, 681–694.

Gómez-Diago, G. (2022). Perspectives to address artificial intelligence in journalism teaching. Revista Latina de Comunicación Social, 80, 29–45.

Hanna, M, & Dodd, M. (2020). McNae's Essential Law for Journalists (25th edition). OUP Oxford.

Hannaford, L. (2015). Computational journalism in the UK newsroom. Journalism Education, 4.

Hansen, A, & Machin, D. (2019). Media and communication research methods (Second edition). Macmillan International Higher Education.

Hughes, A. (2023). ChatGPT: Everything you need to know about OpenAI's GPT-3 tool. BBC Science Focus Magazine. https://www.sciencefocus.com/future-technology/gpt-3/.

Juntunen, L. (2010). Explaining the Need for Speed. In Justin Lewis & Stephen Cushion (Eds.), Has 24-hour News Changed the World? Peter Lang.

Kirk, A. (2016). Data Visualisation: A Handbook for Data Driven Design (First Edition). SAGE.

Komatsu, Tomoko, Gutierrez Lopez, Marisela, Makri, Stephann, Porlezza, Colin, Cooper, Glenda, MacFarlane, Andrew, & Missaoui, Sondess. (2020). AI should embody our values. Proceedings of the 11th Nordic Conference on Human-Computer Interaction, 1–13.

Kothari, A, & Hickerson, A. (2020). Challenges for journalism education in the era of automation. Media Practice and Education, 21(3), 212–228.

Lee, Y. (2022). Beyond online search strategies. Journal of Computer Assisted Learning, 38(4), 1102–1114.

Matsiola, M, Spiliopoulos, P, Kotsakis, R, Nicolaou, C, & Podara, A. (2019). Technology-enhanced learning in audiovisual education. Education Sciences, 9(1), 62.

Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution. https://www.ssoar.info/ssoar/handle/document/39517

Mayson, S. G. (2019). Bias in, bias out. The Yale Law Journal, 2218–2300.

McCarthy, J. (2016, March 9). Mark Ruffalo and Leonardo DiCaprio bring fight against fracking to Los Angeles. Global Citizen. https://www.globalcitizen.org/en/content/mark-ruffalo-and-leonardo-dicaprio-bring-fight-aga/.

Minkin, L. (1997). Exits and Entrances: Political Research as a Creative Art (Vol. 1). Sheffield Hallam University Press.

Miroshnichenko, A. (2018). AI to Bypass Creativity. Information, 9(7), Article 7.

Montal, T, & Reich, Z. (2017). I, robot. You, journalist. Who is the author? Digital Journalism, 5(7), 829–849.

Noble, S.U. (2018). Algorithms of Oppression. NYU Press.

Nünning, V. (2015). Unreliable Narration and Trustworthiness. De Gruyter.

Nygren, T, & Guath, M. (2019). Swedish teenagers' difficulties and abilities to determine digital news credibility. Nordicom Review, 40(1), 23–42.

Nygren, T, & Guath, M. (2022). Students Evaluating and Corroborating Digital News. Scandinavian Journal of Educational Research, 66(4), 549–565.

Ombelet, P, Kuczerawy, A, & Valcke, P. (2016). Employing Robot Journalists: Legal Implications, Considerations and Recommendations. Proceedings of the 25th International Conference Companion on World Wide Web, 731–736.

Pavlik, J. V. (2023). Collaborating With ChatGPT. Journalism & Mass Communication Educator, 78(1), 84–93.

Pinto-Martinho, A., Cardoso, G., & Crespo, M. (2022). AI and journalism, robot journalism and algorithms. In New skills for journalists (pp. 157–169). Transylvanian Museum Society.

Prensky, M. (2001). Digital natives, digital immigrants part 2: Do they really think differently? On the Horizon.

Purington, A, Taft, Jessie G., Sannon, S., Bazarova, N. N., & Taylor, S.T.. (2017). "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 2853–2859.

Simon, F.M. (2022). Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy. Digital Journalism, 10(10), 1832–1854.

Sundar, S. S., & Liao, M. (2023). Calling BS on ChatGPT. Journalism & Communication Monographs, 25(2), 165–180.

Thurman, N., Dörr, K., & Kunert, J. (2017). When Reporters Get Hands-on with Robo-Writing. Digital Journalism, 5(10), 1240–1259.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., & Zoph, B. (2022). Emergent Abilities of Large Language Models (arXiv:2206.07682). arXiv.

Wineburg, S., & McGrew, S. (2017). Lateral Reading: Reading Less and Learning More When Evaluating Digital Information (SSRN Scholarly Paper 3048994).

Acknowledgements: The authors would like to thank all of those who took part in the study. Without their participation, insights and observations, this paper would not exist.

REFERENCE

Sean Tunney, Adam Cox, Athanasia Batziou and Yuwei Lin (2023) 'AI in the newsroom: implications for educators from an experiment with trainee journalists' in *Journalism Education* 12(1) pp 36-46