City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

How to Blend Journalistic Expertise with Artificial Intelligence for Research and Verifying News Stories?

Sondess Missaoui, Marisela Gutierrez-Lopez, Andrew MacFarlane, Stephann Makri
Centre for HCI Design
City, University of London
London, England
name.lastname@city.ac.uk

Colin Porlezza, Glenda Cooper
Department of Journalism
City, University of London
London, England
name.lastname@city.ac.uk

ABSTRACT

The use of AI technology can help to automate news verification workflows, while significantly innovating journalism practices. However, most existing systems are designed in isolation without interactive collaboration with journalists. 'DMINR' project aims to bring humans-at-the-center of AI loop for developing a powerful tool that is sympathetic to the way journalists work. In this paper, we attempt to understand how AI can shape journalists' practices and, crucially, be shaped by them; we aim to design human-centred AI tool that works in synergy with journalists' practices and strike a useful balance between human and machine intelligence. In this paper, we conducted a Co-design workshop to inform the design of the 'DMINR' system. Based on the findings, we outline the main challenges for designing AI systems in the context of journalism, that can serve as a resource for Human-AI interaction design.

KEYWORDS

AI; News verification;Information access; Human-in-the-loop; Journalism; Bias; Transparency.

INTRODUCTION

The spread of misinformation, makes it more difficult for journalists to verify information sources [2]. Finding newsworthiness in complex and big data sets is time consuming and highly difficult process;

Figure 1: Card game to explore journalistic workflow and sources of input.

Conversation starter: "fill the blank" activity to surface assumptions and desires about technology.

Card game 1: deck of 21 cards to explore journalistic practices and values. Journalists collaboratively described some of the journalistic practices including the information verification. Using a card game, the journalists described the investigative reporting process, e.g., how to find a story lead, how to gather evidence, how to explore information and datasets, etc.

Card game 2: deck of 14 cards to provoke discussion around AI technologies. Journalists used statements cards that include statements about AI myths and opportunities in journalism context. These cards help the participants to reflect better about the future of journalism in digital era. Journalists expressed their opinions about the role of AI in supporting or hindering their journalistic practices.

Sidebar 1: DMINR co-design workshop activities

hence, there exists a growing demand for efficient human-centred AI tools that can support journalists in researching and verifying information for news stories creation. The *Google News Initiative*[1] (DNI) funded R&D project DMINR[2] aims to respond to some issues in the area of investigative journalism. In an attempt to help journalists making meaningful connections between data sources, DMINR AI system, which is under development, aggregates data from different reliable sources, merging and monitoring the data to assist journalists in the news verification and production process. In particular, we leverage information retrieval and deep learning technologies to reliably gather and analyze big data from various public sources, extracting and aggregating latent patterns and trends from them. Designing a Human-centred *Artificial Intelligence* (AI) system that is sympathetic to journalistic expertise and values is on the focus of our project. While most existing AI systems in the journalistic domain [6, 8] are developed opaquely and in isolation from the users [7], the DMINR project aims to develop a tool with the journalists involvement throughout the design process. As experts in Artificial Intelligence, we are working together with a group of journalists to build a Human-centred AI tool that work in synergy with journalists' practices and strike a useful balance between human and machine intelligence. Through the *Human-in-the-Loop* (HITL) approach [11], that requires users interaction, we are opening space to discuss and study the challenges in designing an AI tool that is sympathetic to the way journalists work. In the following, we will discuss opportunities and challenges for designing Human-centred AI in the journalism context.

INTERNAL DMINR CO-DESIGN WORKSHOP

We employed a Co-design approach to actively engage the journalists in design activities, equally addressing the following challenges: (*a*) understanding how journalists get the investigative work done in terms of journalistic practices; (*b*) identifying information sources used for a variety of different news areas (business, legal, etc.) for accurate investigative reporting; (*c*) identify the journalists' perceptions toward the role of technology in supporting or hindering journalistic practices. We ran the first internal Co-design workshop for two hours with the participation of seven DMINR team members: four expert journalists, who have decades of experience in newsrooms, one HCI expert as attendees. An information retrieval researcher and a design researcher acted as facilitators (first and second authors respectively). The workshop was clearly differentiated from a "regular" project meeting: we set up a dedicated space and time, crafted original material for the session, and recorded audio with consent of the attendees. Except for the facilitators, the attendees were blind to the specific methodology up until the actual session. The workshop was organized in three design activities, briefly described in Sidebar 1. We focused on discussing journalistic concepts through card games as an approachable, playful way to generate and communicate ideas [9]. The content of the decks of cards (see Figure 1 and Sidebar 1) was based on our previous engagements with journalists and

Q1: *"so if you have an AI that, puts* **all the information**, **all the data** *in a logical way […]".*
Q2: *"the processing power of AI in terms of what it can do in* **bringing an infinite number of sources together***."* '
- Journalist 1 (J1)

Q3: *"AI would [NOT] replace journalists! It won't and it won't because it is a creative process".*
Q4: *"algorithms are too opaque."* - Journalist 2 (J2)

Q5: *"The story angle is everything to the journalist[…], in journalism what is important is the human interest, the interviewing, and the impact of the fact checking on humans. So journalism in my mind is actually talking to people, and AI can't do that."*
Q6: *"I am wary that AI would tarnish the integrity of news by making sources too opaque."* - Journalist 3 (J3)
Q7: *"I said, as a journalist I concerned that AI will generate unintended or maligned consequences…It could be* **bias***, yes, that could be one malign consequence. Or it could be simply, doesn't work as intended because* **the user would not understand how it works***. That can be the case.*
Q8: *"when something is out of my control, it feels as a bigger thing, as a journalist I feel* **protective***."* - Journalist 4 (J4)

Sidebar 2: Quotations from internal workshop participants.

a literature review. The data gathered was transcribed, anonymised, and analysed using Thematic Analysis [3].

JOURNALISTS PERCEPTIONS TOWARD AI

The co-design workshop's results helped us to identify some crucial considerations to be made during the design process of DMINR tool. At an early stage, we noticed three main views that journalists held surround "the automatization" and what AI can and cannot do. Firstly, the discussion revealed that journalists generally overestimate the capabilities of AI when focusing on the potential opportunities afforded by the technology - as seen in the statement by journalist J1 elicited the first activity (see Q1, Q2, sidebar2) . S/He referred to the AI ability to bring an infinite number of data sources automatically. There is no question that AI has the potential to analyze big data from multiple sources quickly and easily, however, it is important to not overestimate the the utility of of Artificial Intelligence in its own right. In fact, to sift through the right data, AI requires journalists' expertise to determine the reliable data sources for their reporting. Hence, an appropriate involvement of the journalists in the design process would be crucial to gain into the right system input. Aside from this relatively optimistic view of AI technology, two further attitudes of journalists toward AI capabilities has been identified in the analysis. Journalists 2 and 3 (J2 and J3) believe that investigative reporting practices and journalism, in general, are beyond AI's capabilities [4] (see Q3, Q5, sidebar2). Our participants declared that the journalism is a creative human practice that cannot be automatized. They placed more weight on the limitations of AI, which indicates a conservative attitude toward its use in the journalism. According to Shoemaker and Reese, 2014 [10], journalists continue to enjoy the status of "sole news story producers' ", which enable them to act as impactful actors influencing the content of news. Thereby, engaging the journalists in a Human-AI interaction design opens space to discuss AI's aims; augment journalists knowledge with meaningful information and subtle pattern, not to replace them [5].

Finally, we found that some journalists adopt a relatively positive blueprint for AI despite the recognition of certain threats. Journalists expressed a concern that AI may hinder the verification practices because of *bias* and *non-transparency* issues. One of our participants, who is a world-renowned expert on journalism (J4), expressed his disquiet about AI bias and lack of *Transparency* (see Q5, sidebar 2). The other participants agreed with him and stressed more the problem of algorithm opacity (see Q6, Q7 sidebar 2). The discussion revealed that the system's *Transparency* could impact on practices and ethical standards in journalism [1]. Journalists suggest that establishing new standards of data sources transparency could help improving reporting "accuracy". A clear concern about system's *Transparency* was revealed by a journalist who discussed the potential that lack of *Transparency* and *Control* could have on the journalists trust and attitude toward a new technology. The journalist claimed that s/he could resist to any technology if it is out of her/his control (see Q8, sidebar 2). In the

following discussion, we will clarify the discovered concepts that present challenges for developing the DMINR tool.

DISCUSSION AND GUIDANCE

Although our workshop was small, the participation of journalists, with several decades of newsrooms experience, provides preliminary results of potential interest toward the design of DMINR tool. The following are the key findings from our co-desing workshop:

- **The Journalists' perception of AI is "broad" but not "deep".** Our participants had a broad knowledge of various aspects of AI. They were able to highlight certain AI abilities, e.g., analyzing big data and extracting meaningful insights from it, but they were less clear on how algorithms worked and AI limitations. This has raised a lack of trust toward AI which they consider as 'black box'.
- **Lack of explanation leads to high expectations**. AI technology in general lacks of explanation. In the journalism context, the lack of explanation raised many challenges including high expectations. Some journalists claim that AI could perform tasks that are currently beyond the state-of-the-art in journalism, e.g., sifting through an infinite amount of data sources with no limits. Some of their thoughts reveal how various errors and malfeasances in explaining AI have created overestimation about practical AI capabilities.
- **Transparency and Control are interlinked concepts.** Journalists have raised issues about their trust in the AI systems. They estimate that AI would tarnish the integrity of news by making sources too opaque. This implies that *Transparency* is paramount for any AI technology in the journalism context. In addition, journalists expressed their disquiet about how AI monitors the data. The "automatic filtering" of the data are risky in several ways for them, i.e., missing out important information, lack of opportunity, etc. In relation thereto, journalists required to gain *control* over the system to protect journalistic ethics and standards. Given that, *Transparency* and *Control* are interlinked concepts that: (i) provide insights into the AI system's vulnerability; and (ii) enable the system to be more verifiable and trustworthy;

CONCLUSION

Regardless of whether a system best supports journalists' in their practices, its actual performance is subject to journalists' attitudes towards the technology used. Indeed, journalists' attitudes towards AI could influence the success or failure of any tool even if it could positively impacting their work. We need to be engaged with journalists in discussion about design to: (a) to understand how AI can shape their practices and, crucially, be shaped by them; (b) design a human-centred AI technologies that are sympathetic to the way they work.

REFERENCES

[1] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 3 (2018), 973–989.

[2] Vian Bakir and Andrew McStay. 2018. Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism* 6, 2 (2018), 154–175.

[3] Virginia Braun and Victoria Clarke. 2013. *Successful Qualitative Research: A Practical Guide for Beginners*. SAGE Publications Ltd, London.

[4] Meredith Broussard. 2015. Artificial intelligence for investigative reporting: Using an expert system to enhance journalist-sâĂŹ ability to discover original public affairs stories. *Digital Journalism* 3, 6 (2015), 814–831.

[5] Matt Carlson. 2015. The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital journalism* 3, 3 (2015), 416–431.

[6] Qi Fan, Yuchen Li, Dongxiang Zhang, and Kian-Lee Tan. 2017. Discovering newsworthy themes from sequenced data: A step towards computational journalism. *IEEE Transactions on Knowledge and Data Engineering* 29, 7 (2017), 1398.

[7] Tarleton Gillespie and Nick Seaver. 2016. Critical algorithm studies: A reading list. *Social Media Collective* (2016).

[8] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. 2017. Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1803–1812.

[9] Andres Lucero and Juha Arrasvuori. 2010. PLEX Cards : A Source of Inspiration when Designing for Playfulness. In *Fun and Games 2010*. 28–37.

[10] Pamela J Shoemaker and Stephen D Reese. 2013. *Mediating the message in the 21st century: A media sociology perspective*. Routledge.

[11] Fabio Massimo Zanzotto. 2017. Human-in-the-loop Artificial Intelligence. *arXiv preprint arXiv:1710.08191* (2017).