



Latest updates: <https://dl.acm.org/doi/10.1145/3630106.3658987>

RESEARCH-ARTICLE

## The Impact and Opportunities of Generative AI in Fact-Checking

**ROBERT WOLFE**, University of Washington, Seattle, WA, United States

**TANUSHREE MITRA**, University of Washington, Seattle, WA, United States

Open Access Support provided by:

University of Washington



PDF Download  
3630106.3658987.pdf  
29 December 2025  
Total Citations: 19  
Total Downloads: 3667

Published: 03 June 2024

Citation in BibTeX format

FAccT '24: The 2024 ACM Conference  
on Fairness, Accountability, and  
Transparency

June 3 - 6, 2024  
Rio de Janeiro, Brazil

# The Impact and Opportunities of Generative AI in Fact-Checking

Robert Wolfe  
University of Washington  
Seattle, Washington, United States  
rwolfe3@uw.edu

Tanushree Mitra  
University of Washington  
Seattle, Washington, United States  
rwolfe3@uw.edu

## ABSTRACT

Generative AI appears poised to transform white collar professions, with more than 90% of Fortune 500 companies using OpenAI's flagship GPT models, which have been characterized as "general purpose technologies" capable of effecting epochal changes in the economy. But how will such technologies impact organizations whose job is to verify and report factual information, and to ensure the health of the information ecosystem? To investigate this question, we conducted 30 interviews with  $N=38$  participants working at 29 fact-checking organizations across six continents, asking about how they use generative AI and the opportunities and challenges they see in the technology. We found that uses of generative AI envisioned by fact-checkers differ based on organizational infrastructure, with applications for quality assurance in Editing, for trend analysis in Investigation, and for information literacy in Advocacy. We used the TOE framework to describe participant concerns ranging from the Technological (lack of transparency), to the Organizational (resource constraints), to the Environmental (uncertain and evolving policy). Building on the insights of our participants, we describe value tensions between fact-checking and generative AI, and propose a novel Verification dimension to the design space of generative models for information verification work. Finally, we outline an agenda for fairness, accountability, and transparency research to support the responsible use of generative AI in fact-checking. Throughout, we highlight the importance of human infrastructure and labor in producing verified information in collaboration with AI. We expect that this work will inform not only the scientific literature on fact-checking, but also contribute to understanding of organizational adaptation to a powerful but unreliable new technology.

## CCS CONCEPTS

- Social and professional topics → Computing industry;
- Computing methodologies → Artificial intelligence;
- Human-centered computing;
- Applied computing → Computers in other domains;

## KEYWORDS

Generative AI, Fact-Checking, Transparency, Sociotechnical Infrastructure, Design



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

## ACM Reference Format:

Robert Wolfe and Tanushree Mitra. 2024. The Impact and Opportunities of Generative AI in Fact-Checking. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil*. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3630106.3658987>

## 1 INTRODUCTION

A research report issued by OpenAI in March 2023 [25], days after the release of its flagship GPT-4 model, contended that generative pretrained transformers (GPTs) are general purpose technologies, technologies with the potential to reshape not an individual profession but an entire economy. Unlike many previous general purpose technologies, the authors asserted that generative AI will impact primarily professions with a higher barrier to entry, those requiring more education and experience to carry out. Among the professions estimated by an OpenAI model as "fully exposed" to transformation by generative AI, defined as reducing by at least 50% the time needed to complete the tasks of an occupation, was "News Analysts, Reporters, and Journalists" [25]. Yet these professions have outsized epistemic effects on society [24], as they remain the primary means for producing knowledge claims and for critically assessing sources and information [23], thus ensuring the integrity of the online information space. If generative AI is to reshape such roles, understanding how it might do so – and where to draw the boundaries – is crucial to ensure the health of the information ecosystem.

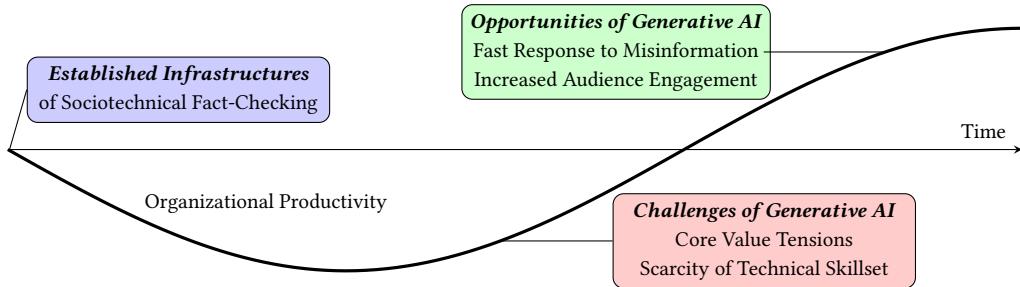
In this work, we study the impact of generative AI in fact-checking, a profession that specializes in determining the reliability of information disseminated through traditional and social media, undertaken at publishing houses and independent organizations around the world. Fact-checking is a complex sociotechnical process, involving human judgment exercised in conjunction with AI-based tools to observe misinforming claims and narratives as they spread [37]. While most fact-checking organizations necessarily embrace technological tools, they are skeptical of technologies that promise to automate large parts of the fact-checking process, and deprioritize or displace human expertise [41]. Understanding perspectives of key stakeholders at fact-checking organizations is thus important to facilitate adoption of a technology that could help respond efficiently to misinformation, while prioritizing the role of human expertise. We address two research questions:

### (1) RQ1: Opportunities of Generative AI in Fact-Checking:

What opportunities do fact-checking organizations see in generative AI? How are organizations presently using generative AI, and how do they envision using it?

### (2) RQ2: Challenges and Limitations of Generative AI in Fact-Checking:

What challenges do fact-checkers see in using generative AI to support their work? What prevents them from further incorporating generative models?



**Figure 1:** Incorporating a society-altering technology like generative AI into sociotechnical fact-checking work requires “intangible” investments [9] (new processes and skills) to realize its potential without deprioritizing the values of fact-checking or displacing the role of human experts.

To address these questions, we interviewed  $N=38$  participants at 29 fact-checking organizations in a range of roles, from investigation, to management, to engineering. We captured diverse, global perspectives from participants located across 19 countries and six continents. Interviews provided detailed accounts of where fact-checkers envisioned using generative AI, and concrete examples of applications in use or in development. Participants also shared barriers to adopting generative AI, ranging from technical limitations to value misalignments. Figure 1 draws on organizational research [9] to illustrate the investment to overcome these challenges and realize the benefits envisioned by participants. We make four contributions:

- **Enumerating Opportunities and Limitations of Generative AI in Fact-Checking:** We describe the opportunities for generative AI in five fact-checking infrastructures (Editing, Investigation, Audience Management, Technology, and Advocacy), and adopt the Technology-Organization-Environment framework [62] to describe challenges.
- **Designing for Verification:** We propose a novel dimension in the design space for generative models that centers Verification, or ensuring the veracity of content. We describe this dimension with a  $2 \times 2$  matrix, with the Producer of content on the X axis, and its Verifier on the Y axis, and discuss its use beyond fact-checking in high stakes domains.
- **Mapping Value Tensions:** Using the principles of the International Fact Checking Network (IFCN) [60] as a basis for the sociotechnical values of fact-checking, we describe value tensions between fact-checking, which centers transparency and reliability, and generative AI, a technology exhibiting unpredictable and often unreliable behavior.
- **Defining a Research Agenda:** We propose nine directions for fairness, accountability, and transparency researchers to develop technologies, designs, and approaches supporting responsible use of generative AI in fact-checking.

## 2 RELATED WORK

**Sociotechnical Infrastructures of Fact-Checking.** “Fact-checking” refers to the investigation of potentially misinforming claims and narratives that may adversely impact individuals and society [34, 36]. Fact-checking is primarily a “socio-technical” task [70, 90], wherein technology is useful and meaningful only in the context of its relationship to the humans who interact with it [81, 92]. While fact-checkers necessarily employ data-driven technologies [22, 37],

and envision further uses of technologies to, for example, minimize the amount of harmful content to which they are exposed [41], human judgment is also crucial to the fact-checking process [35], and fact-checkers are skeptical of technologies that promise to fully automate parts of fact-checking work [41]. Prior work has sought to clarify the communities [7] and sociotechnical infrastructure undergirding the processes of fact-checking. Juneja and Mitra [41] describes fact-checking organizations as composed of “human and algorithmic infrastructures” fulfilling distinct roles in fact-checking, such as editing and investigation. We draw on these roles when unpacking the opportunities presented by generative AI in fact-checking.

**Organizational Change.** Eloundou et al. [25] contend that generative AI is a “general purpose technology” that will reshape society and the economy, with impacts greater for professions requiring high levels of education. Brynjolfsson et al. [9] describes general purpose technologies as necessitating intangible “complementary investments” to realize their potential, such as “co-invention of new processes, products, business models and human capital,” suggesting the sociotechnical nature of technology adoption. Prasad Agrawal [62] describe generative AI’s adoption using the Technology - Organization - Environment (TOE) framework, noting the impact of factors like regulation and organization size. In a now foundational text, Fichman and Kemerer [30] note that the widespread acquisition of a technology by organizations may not result in its widespread deployment, especially where “knowledge barriers” mitigate effective use. We borrow concepts from the TOE framework to tease apart the challenges faced by fact-checkers in adoption of this new technology.

**Generative AI.** Generative AI technologies such as ChatGPT [55] and its predecessors [8, 68, 69] can ingest human input in natural language [56] and, depending on their architecture and training objective, produce palatable text [40, 85, 86], images [71, 72, 76], video [77, 88], and source code [14, 89]. Such technologies both pose difficulties for fact-checkers, who must contend with higher quality misinformation produced more easily [39, 42, 91], but also opportunities for novel technologies in their work [19, 75]. Recent work highlights difficulties with generative AI for journalism and fact-checking, including low audience trust in AI-generated content [46] and biases in the dissemination of AI-assisted fact-checks [53]. Researchers in HCI have mapped the design space [10] of generative AI [51], describing interactions possible with users and ways to use it in

domains like scientific research [50] and creative writing [12]. Building on participatory design [5, 79], recent work develops “participatory AI,” wherein human subjects envision new AI-driven designs with researchers [3, 20, 66]. For fact-checking, Das et al. [19] conduct a review of human-centered NLP and develop a confusion matrix for calibrating trust in human-AI collaborations. We extend this line of work by proposing a socio-technical verification dimension.

**Values in AI and Fact-Checking.** Scholars have found that AI and machine learning research is not “value neutral” but prioritizes values like performance and generalization, while neglecting considerations like “negative potential” [4]. Bender et al. [2] contend that training models on poorly specified textual data poses numerous ethical risks for downstream use. Fact-checkers adhere to rigorous ethical codes [57], such as those set forth by coalitions like the 172-member (as of this writing) International Fact Checking Network (IFCN) [60], which publishes a code of principles to which its signatories commit [61]. We use the published principles of the IFCN, an organization formed to promote common standards in fact-checking [41], to explore value tensions between generative AI and fact-checking.

### 3 METHODS

#### 3.1 Participant Recruitment

We conducted an interview study with  $N=38$  employees of fact-checking organizations or teams in publication houses, with experience in their current role ranging from 1 year to 18 years. As shown in Table 1, we recruited from a total of 29 fact-checking organizations using purposive sampling and snowball sampling [26, 52], first reaching out to potential participants by sending an email advertising the study to the listserv of the IFCN. This resulted in three interviews with six participants working at three organizations. The Community Manager of the IFCN then provided us with contact information for six potential participants for our study, to whom we reached out. This resulted in three interviews with three participants at three organizations. We next utilized a list of 23 fact-checkers known to one of the authors, who has maintained a long-term relationship with the global fact-checking community. This resulted in five interviews with six participants at five organizations. Finally, we sent cold emails to 60 IFCN signatory organizations, explaining our interest in an interview and how we found their contact information. We recruited in this way not only to increase the number of participants in the study, but also to increase the study’s global reach, as we emailed primarily organizations in developing countries and the global south. This strategy resulted in 14 interviews with 18 participants at 14 organizations across five continents. We employed snowball sampling when participants offered to connect us with a participant well-suited to the study, and reached out via email. This resulted in five interviews with five individuals working at five organizations.

#### 3.2 Interview Protocol

We created a semi-structured interview protocol that posed general questions regarding the use and impact of generative AI in fact-checking. We began interviews by asking participants to tell us about their background, including their position, experience in fact-checking, and familiarity with generative AI. We asked about their

company’s background, including the size and technical experience of the fact-checking team, and how long the company had been performing fact-checking work. We then explicitly posed the primary research questions of our study, asking participants to characterize 1) how they used generative AI in their work; 2) opportunities for using generative AI in fact-checking; 3) challenges and limitations of using generative AI in fact-checking; and 4) how researchers could design generative technologies that better support fact-checkers. We asked participants to clarify, discuss, and expand upon responses to better understand their perspectives. We also asked follow up questions where appropriate about several specific topics, including the use of corporate vs. open source AI; modalities (text, image, etc.) of misinformation they use generative AI to handle; use of generative AI to handle narratives; guidelines for using generative AI in their organization; and impacts the participant witnessed generative AI having. We submitted the interview protocol as part of the supporting materials to our University’s Institutional Review Board.

#### 3.3 Interview Process

We conducted 30 interviews between October 2023 and January 2024. Multiple participants attended seven interviews, with one participant typically a manager, and the other(s) involved in technology or investigation. In one case, we interviewed two managers who passed follow-up questions to the engineering team, forwarding their responses to us by email. Interviews lasted 30 to 90 minutes, averaging approximately 45 minutes. Interviews were conducted solely in English. We accommodated the request of one participant to send answers by email because they preferred writing over speaking in English. The participant then also met with the first author for 20 minutes. Participants who used generative AI sometimes shared their screen and displayed the interfaces used with these technologies. One participant shared a Jupyter notebook showing their use of AI in a data science pipeline. Other participants linked us to Github pages or company technical reports. We did not offer to compensate participants to prevent feelings of coercion.

#### 3.4 Data Analysis

After transcribing the interviews, we deductively coded them according to four categories tied closely to the research questions: Present Use of Generative AI in Fact-Checking; Opportunities to Use Generative AI in Fact-Checking; Challenges and Limitations to Using Generative AI in Fact-Checking; and Ways Computational Research in Fairness and Transparency Can Support Fact-Checking. We then conducted inductive coding within each deductive category.

The authors first coded four interview transcripts, after which the first author created a codebook that included inductively derived codes organized within the deductive categories. The codebook consisted of the names of codes, explanations of the codes, and the associated participant quotes. The first author shared the codebook, and the last author offered feedback and suggestions, after which the authors met to discuss the codes and revised or removed codes on which they could not reach agreement. The authors then coded four additional transcripts at a time, updating the codebook after each round with more precise definitions and additional context

Continents (6)   Countries (19)	Fact-checking Organizations (29)
	Australian Associated Press [64], Agence France-Presse [65], Africa Check [13], Aos Fatos [28], Chequeado [15], Code for Africa [31], ColombiaCheck [17], Der Spiegel [78], Factly [27], India Today [84], Infoveritas [38], Lead Stories [80], Litmus [44], logically.ai [45], Maldita.es [47], Meedan [48], MindaNews [49], Newtral [54], Pagella Politica [58], PolitiFact [59], Pravda [63], Rappler [73], RMIT FactLab CrossCheck [18], Science Feedback [29], Taiwan FactCheck Center [11], Tech4Peace [82], The Quint [67], Thomson Reuters [74], Univision El Detector [21]

Table 1: Participants were recruited from 6 continents, 19 countries, and 29 fact-checking organizations.

from participant quotes. Next, the authors followed a thematic analysis process [6, 16] to generate themes that described the findings and addressed the research questions. Specifically, the authors reviewed the codes and their associated participant quotes, drafted memos describing proposed themes, met to discuss the proposed themes, and converged on a set of final themes on which agreement could be reached. These themes form the basis of the Findings section.

## 4 FINDINGS: OPPORTUNITIES AND CHALLENGES OF GENERATIVE AI IN FACT-CHECKING

### 4.1 RQ1: Opportunities of Generative AI

We found during thematic analysis that the technologies used and envisioned by fact-checking organizations depend largely on the organizational infrastructures into which they would be integrated. These infrastructures accorded to a large degree with the work of Juneja and Mitra [41], which described the sociotechnical work of editors; fact-checkers; social media managers; and long-term advocates. Drawing on these roles, we organize our findings according to the following divisions of organizational infrastructure, as shown in Fig. 2: Editing, Investigation, Audience Management, Advocacy, and Technology, the last of which is new to our work but necessary to describe the impact of generative AI on the work of software developers and data scientists who build and maintain the data science pipelines employed by fact-checking organizations.

When describing a technology, we also identify its status according to three levels of maturity, denoted using icons:

- **► In Use:** Technologies presently in use by participants, denoted with a rightward arrow to evoke a “Play” symbol.
- **☒ In Progress:** Technologies undergoing prototyping, testing, betas, or development by our participants, denoted with an hourglass symbol to communicate that some time remains before these technologies will be implemented.
- **👓 Envisioned:** Technologies envisioned but not implemented or prototyped by our participants, denoted with eyeglasses to communicate that these technologies are further away and not yet in development.

Where fact-checking organizations reported achieving differing levels of maturity for a technology, we describe the most mature, and make note if this level of maturity has not been achieved by most other fact-checking organizations.

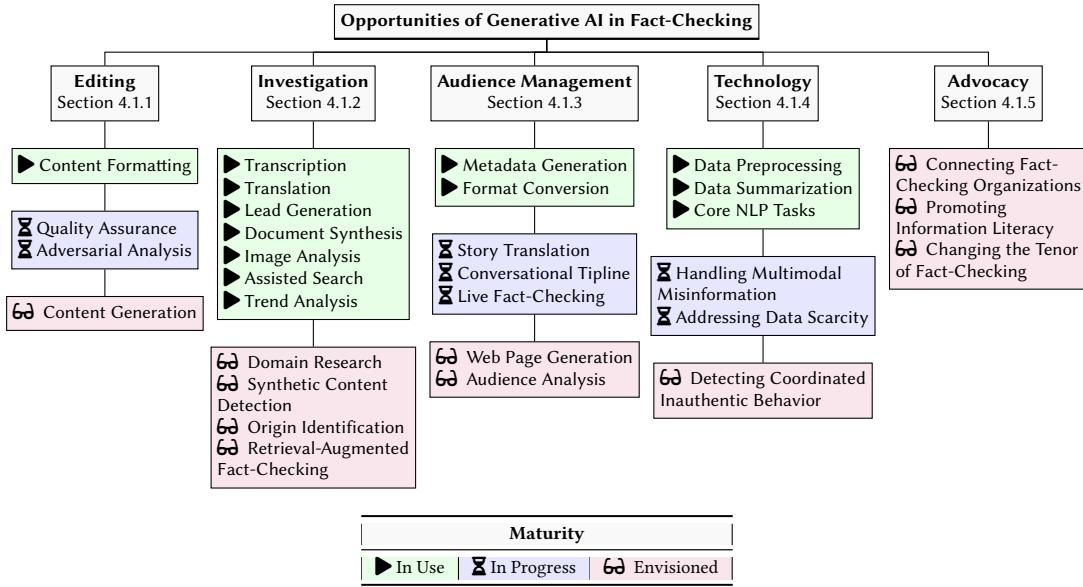
**4.1.1 Editing.** Editing ensures fact-checking content is engaging, approachable, and error-free, and spans from the beginning of a fact-check to publication, as editors are often involved in deciding which claims are worthy of a fact-check [41]. Participants in our study who were involved in Editing usually reported managing small teams.

Many participants described using generative AI to refine and restructure fact-checks and internal reports, which then undergo human review. P3 reported using “premium” ChatGPT for ► **Content Formatting** - to edit and refine written reports, and to restructure “dense content” into an approachable format for readers. P18 noted using ChatGPT to help in “brushing up text.” Several participants described in-development applications of generative AI for systematic ☒ **Quality Assurance**, to prevent cosmetic and substantive editorial errors. P35 reported using ChatGPT to highlight grammatical and factual errors, tasks for which they normally use Grammarly [33]. P27 developed an app to address such errors:

We began with simple mistakes, geographical errors, or misspellings [...] I made a little Shiny app out of [ChatGPT] and showed it around, and people were really like [...] I can really see how this helps us. [...] And my aim is that we don’t have these kind of simple mistakes anymore [...] this would be a huge achievement because simple mistakes are trivial on the one hand, but on the other hand [...] it’s really important for the sense of quality the reader has and for trust.

P3 expanded on this view and envisioned generative AI providing ☒ **Adversarial Analysis** before publishing a fact-check, referencing a strategy in development at a Sudanese newsroom: “They use generative AI to actually give it an article that has been written, and ask the model to actually tell us whether there are any assumptions that have been met, that are inaccurate or incorrect.” Finally, P8 envisioned fine-tuning a generative model on verified fact-checking articles, and using it for **👓 Content Generation** at the beginning of the composition process, improving “fact-checking output [by] pre-writing those fact-check articles,” provided that the content did not require deep scientific knowledge of a topic, and remained subject to human review. While some participants envisioned using generative AI for content generation, many expressed deep discomfort with generative AI writing fact-checks, a finding explored in sections 4.2.1 and 5.1.2.

**4.1.2 Investigation.** Investigation refers to the process of assessing the accuracy of potentially misinforming claims, and involves monitoring online sources for misinforming content; gathering



**Figure 2: A description of In Use, In Progress, and Envisioned generative technologies grouped according to five fact-checking infrastructures.**

verified sources to substantiate or refute claims; and writing a fact-check or internal report [41]. 19 of our 38 participants described working primarily in Investigation.

Participants described many ongoing and envisioned uses of generative AI to perform tasks related to investigation and research assistance. As P25 noted, generative AI is used in common and “taken-for-granted” tasks like ► **Transcription**, which save time and money for fact-checkers. P20 highlighted the use of generative AI for ► **Translation** of internet content in need of investigation, such as “many translations from Ukraine” due to misinformation related to the Russia-Ukraine war. Participants also described adopting generative AI to directly support investigations. P11 noted that, while they never use ChatGPT for writing fact-checks, they use it for ► **Lead Generation**, “trying to generate ideas for stories.” P28 used ChatGPT for ► **Document Synthesis**, to save time by organizing research notes and summarizing text from web pages. P19 used GPT-4 for ► **Image Analysis**, substituting the model for reverse image search in some cases, noting they can “ask it where the photo was taken, and sometimes we [get the] correct answer,” or useful hints for continuing the search. P25 fine-tuned GPT-3.5-Turbo to perform ► **Assisted Search**, generating custom Google search queries, often in a language not spoken by the fact-checker, noting that such a task would “take me hours to do and I still might miss some of the terms.” P14 reported using ChatGPT for ► **Trend Analysis**, to keep abreast of media produced by websites known to produce misinforming content: “We take the top 200 headlines from the last 24 hours from those sites [...] and run them through ChatGPT, asking it to summarize the main narratives [...] and extract the names of people, places, entities [...] and then send that to me by email. So every six hours, [...] we get an email.”

Participants envisioned technologies to increase their expertise and verify novel content. P29 noted fact-checkers need to “become a little mini expert in a certain specific topic,” and envisioned a

technology for ► **Domain Research**, summarizing and collating literature for review. P24 envisioned using generative AI for ► **Synthetic Content Detection**, identifying content produced by AI. P24 described this as their “holy grail” given recent increases in synthetic content and the difficulty of fact-checking it. P29 envisioned a tool for ► **Origin Identification**, scanning the internet for the first occurrence of content, bringing fact-checkers “closer to the verification.” Finally, P26 envisioned a ► **Retrieval-Augmented Fact-Checking** system that could “retrieve data in almost real time, to consult with databases.”

**4.1.3 Audience Management.** Audience management refers to processes supporting the publication and wide dissemination of fact-checking content. Audience managers increase engagement by employing SEO optimization, online advertising, and conversion of written content to short videos [41]. Our interviews revealed that audience management involves connecting with consumers of fact-checks over many channels, of which social media is one.

Participants used generative AI to both connect with existing audiences and reach new audiences. P33 used generative AI for ► **Metadata Generation** to support social media content, including “summaries, SEO for article publishing, title generation.” P14 used generative text-to-speech models for ► **Format Conversion**, taking fact-checks and converting them into audio for short videos posted to “Tiktok, Instagram, YouTube shorts,” noting that AI helps achieve the right volume for disseminating content via video sharing algorithms. P31 described working on AI-based ► **Story Translation** of their fact-checked content into multiple languages, a goal echoed by P14, who envisioned translating their short videos.

Participants also described efforts to connect immediately with audiences, before misinformation could go viral. P32 described a beta of a fact-checking chatbot in a ► **Conversational Tipline** using OpenAI’s GPT-4, from which they collect circulating misinformation from users and instantaneously deliver information to

audiences who use more private platforms like WhatsApp, rather than Facebook and X. Other participants, including P1 and P17, described in-progress chat-based tiplines via which users can submit suspected misinformation. P13 described an in-progress tool for **☒ Live Fact-Checking** that “can do claim matching while a person is speaking,” allowing for claims to be debunked in real time.

Participants envisioned tools for automatic web content generation and predictive analysis of audience engagement. P28 envisioned automatic **☒ Web Page Generation** that could produce fact-checks “based on social media posts that are verified [...] and then just code the iFrames for us to be able to embed it in our own content,” saving programming labor. Finally, P12 envisioned a system for **☒ Audience Analysis**, providing insight into how audiences would consume a factcheck, and recommending it be presented as a video, an infographic, or a short or long-form article.

**4.1.4 Technology.** Technology refers to work building and maintaining data science pipelines used by fact-checking organizations. While not all organizations have a Technology unit, many uses of generative AI would be invisible without specific reference to the work of software engineers and data scientists employed by fact-checkers and their partners.

Participants described in-progress generative technologies to improve the core functionality and end user experience of fact-checking data science pipelines. P2 used generative AI for **► Data Preprocessing**, to “get a rewriting or a restatement of the claim that’s a bit cleaner” than unprocessed social media content or tipline messages. P8 noted that generative models improve on **► Core NLP Tasks** over finetuned BERT models - “the previous generation” of NLP - including for claim content matching. **► Data Summarization** for human end users was another predominant use reported by our participants. For example, P2 described using generative AI to provide a human-readable summary of clusters of misinforming content, describing “the variety of content” and “how it’s changing over time.” P25 tested the capability of Google’s generative models for natively **☒ Handling Multimodal Misinformation**, wherein the relationship between text and image must be parsed to understand subtle misinformation or hate speech. P2 described ongoing experiments for **☒ Addressing Data Scarcity**, noting that they would “generate pseudo labeled data” where real, “gold standard” data for novel misinformation did not exist, and either “use those labels directly or use them to train a lower cost classifier.” Finally, P16 envisioned a generative model for **☒ Detecting Coordinated Inauthentic Behavior** and influence operations, noting AI might “do more meaningful work on [detecting] coordinated networks, behavior.”

**4.1.5 Advocacy.** Advocacy refers to long-term processes to influence information policy, forge connections between fact-checking organizations, and engage with the public via misinformation literacy campaigns [41]. Participants performing advocacy work were often senior managers who also managed teams of investigators.

Participants suggested generative AI could encourage information literacy, promote relationships between organizations, and improve access to information. P8 envisioned generative AI helping in **☒ Connecting Fact-Checking Organizations**, by standardizing the methods and technologies to combat misinformation in Europe, noting that recent models handle most European languages

well. P12 envisioned generative AI scaling fact-check operations by connecting organizations across Africa: “I think that’s one area in which generative AI can really help. If fact-checkers are working together [...] they can help scale the impact of their fact check to different segmented audiences that they serve [...], whether it’s local language, whatever format.” P21 envisioned generative AI for **☒ Promoting Information Literacy**: “people will have the option to kind of play games with the chatbot that are intended for media literacy on misinformation.” Finally, P30 envisioned generative AI **☒ Changing the Tenor of Fact-Checking**, shifting the way audiences consume fact-checking content:

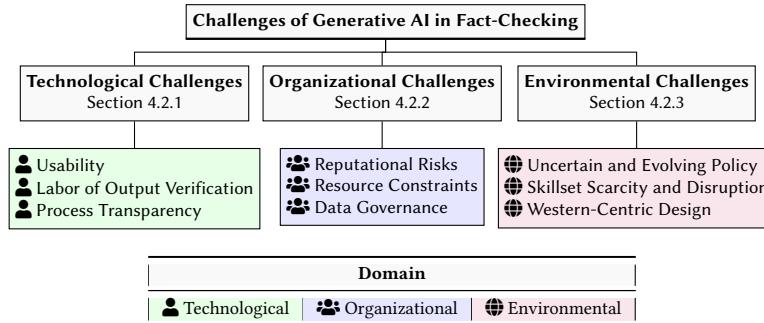
Changing the way users can consume reliable and good information could be incredibly beneficial for fact-checkers [...] If [...] they need and get good information [...] with a chatbot, for example, or with other ways, that would be fantastic. [...] people value us as we are because we give them reliable information and they know they can trust us, but if they also knew that they can consume [that] information in any way they wanted to, I think it would be an incredible leap forward.

## 4.2 RQ2: Challenges and Limitations

We found during thematic analysis that, unlike RQ1, challenges did not break down based on organizational infrastructure. Rather, participants described challenges related to using the technology itself; to incorporating it in an organization; and to factors that affected society as a whole, and were often out of their organization’s control. As illustrated in Fig. 3, the Technology-Organization-Environment (TOE) framework [62] offers a ready model for these findings, and we use it to describe participant challenges as follows: **👤 Technological Challenges** that impact the user of a system, such as the manual labor of verifying generative model outputs, denoted with an icon of a person (note that Technological challenges are in fact sociotechnical, involving human interaction with technology [90]); **👥 Organizational Challenges** that impact an organization in the aggregate, such as reputational risks incurred by using systems that hallucinate, denoted with an icon of multiple people; and **🌐 Environmental Challenges** that impact not only an organization but an entire society, such as the scarcity of skilled workers, denoted with a globe icon.

**4.2.1 **👤 Technological Challenges.**** Participants described barriers related to model usability; the labor of verifying model output; and the conflict in using a hard-to-explain technology in a process requiring absolute transparency.

**👤 Usability:** Participants noted a lack of clarity concerning prompt engineering and hyperparameter tuning. P27 described an iterative process of choosing prompts for OpenAI models that resulted in uncertainty: “We prompted and we coded a bit and we thought, oh, this prompting technique, and combining this prompt with that one and iterating it, and then majority rule. And we [...] thought, okay, is this really the way that this should be used?” P2, who develops AI for fact-checking, noted that generative AI is “incredibly sensitive” to prompts, and “I’m sure we don’t have the best approach” to prompt design. P8 noted the verbosity of ChatGPT reduced its usefulness for fact-checking content, which is “laser-focused.” P14 noted that OpenAI models used for summarization randomly enter



**Figure 3:** A description of the Technological, Organizational, and Environmental challenges to generative AI in fact-checking

“loops” of repeating one word. P25 tweaked settings like temperature to improve reliability, but the effects were hard to see in model output. Finally, P13 expressed frustration over failures of image models like DALL-E to render text in images, limiting their use in creating visual content for stories.

**Labor of Output Verification:** Every participant described human review of AI-generated content as non-negotiable for ensuring the quality of published fact-checking content. Participants described some uses of generative AI as currently untenable due to the verification labor required. P11 summed up participants’ opinions: “Of course it’s not as accurate. The tools are not as accurate. You still need to corroborate the information that you get.” P25 noted that while it is “tempting to use [generative AI] for speeding up your work,” its unreliability means fact-checkers must “see [if] this is correct, what is the source?” P36 noted that, even if ChatGPT provides a lead or answers a question, “it’s just as quick for us to go and find it [...] we’re just so used to that lateral sort of work. And to be honest [...] we’d be going and double checking all that anyway.” P14 noted that, like the internet before it, generative AI re-organized the efforts of fact-checkers to fit the technology, noting “we had to train them” on writing styles that led to better outcomes with generative AI.

**Process Transparency:** Participants described generative AI as a potential impediment to the transparency needed to create trustworthy content. P29 noted, “Our sourcing is [...] always actually quite transparent. [...] we fill our story with hyperlinking to our sources and [...] how we got to everything.” P8 noted that hallucinations prevented them from using ChatGPT: “The result was largely unusable [...] The sources have to be very well integrated [...] it just doesn’t work. Sounds very good, [but] there will be hallucinations, it will just make up sources.” P12 said explaining research that uses generative AI is hard because “as fact-checkers, we actually do not understand the processes” of the models. Finally, P35 noted that generative AI may engender questions about bias concerning selection of experts: “Have [models] weighed whether [...] there are uncertainties about them, have they been disgraced for some reason? [...] humans have biases as well, but I think in factcheck [...] there’s always many, many different sets of eyes on our checks and the experts we use.”

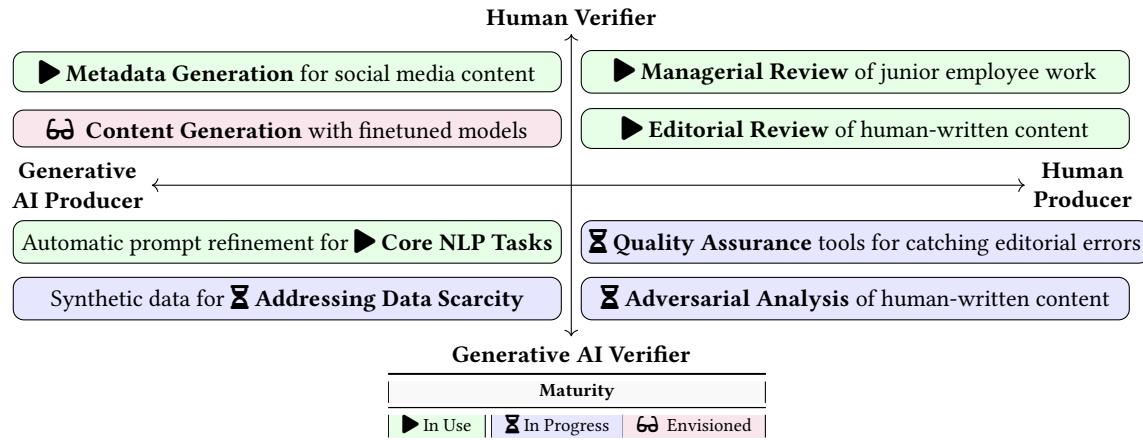
**4.2.2 Organizational Challenges.** Participants described organizational barriers including reputational risks in unpredictable

technology; resource constraints preventing investment; and concerns over data provenance and ownership.

**Reputational Risks:** The most common organizational barrier participants identified was the reputational risk of a mistake in a fact-check. P16 said “just by default we need to be much more cautious than almost anyone else, because it’s hard for us to come back from a big mistake.” P8 and P26 both noted that 90% accuracy is insufficient for fact-checkers, whose relationship with audiences depends on offering information verified by experts. P2 noted that sharing generative technologies with partner organizations also shares risk, prompting further caution: “It’s their organization’s reputation that’s at risk, not ours only.” P18 said they would not trust generative models when there is only one correct answer. Finally, P2 noted that academic evaluations of generative AI are unreliable, as fact-checkers handle *novel* information: “In an academic context, it’s always retrospective. [...] you put that into the Bing API or Google and you find lots of relevant content that can help refute that claim. But when it first appeared [...] I don’t think that was the case.”

**Resource Constraints:** Participants said that most fact-checking organizations lack the financial resources to invest in generative AI. P3 noted that donor funding informs the building of new technologies: “We rely on donor funding a lot, and donor funding is to address a specific use case. So if there’s not enough resources here marked for building a machine learning model, then, we just do that out of pocket or partner with other organizations. So that has been a main limitation for us.” P27 notes that, even at their well-resourced organization, “we don’t have really this AI development department” and that colleagues in low-resource organizations only develop AI tools with universities. P13 said that their organization cannot afford tools developed by better-resourced fact-checkers: “[I] met the team of <Organization> at the last IFCN conference, and they told me it’s going to be a huge sum to get that subscription [...] a small company with seven people, [we] might not be able to afford that.” P21 noted resource constraints facing organizations in the global south:

We’re in the global south, so sometimes the resources are not the same, either to use generative AI [...] for investigating or creating our own tools. For example, some colleagues in Spain have a chatbot [...] and right now we’re trying to find resources to buy the chatbot, the basic form of the chatbot [...] it’s more like an economic problem and it’s not



**Figure 4: Designing for verification: A sociotechnical verification space for the production and verification of content.**

exclusive for [us], but probably more small fact-checking organizations in different global south countries.

**Data Governance:** Participants expressed concerns about the privacy of their data and the provenance of AI training data. P27 noted using open source models when possible, as they “have very sensitive material [...] investigative reporting and investigative stories, and we don’t want this to be used in models and as training material.” P27 appreciated the “legal safety” of European data protection laws, recalling a conversation with a fact-checker who highlighted the importance of trust between organizations: “I really would like to work with all these Google tools, but still, it’s Google and I’m kind of hesitating. I wish the New York Times would’ve developed it. Then it would be very easy for me to trust it.” P31 noted many organizations questioned if they should be compensated for the labor of producing content used to train generative models: “You’ll just continue to see more interest in using fact-check information to feed AI [...] Are we going to be compensated?” P12 described the uncertainty of what data was used to train models as “problematic” for fact-checkers.

**4.2.3 Environmental Challenges.** Participants described society-wide barriers including uncertain government and partner policy; skillset scarcity and disruption; and western-centric design.

**Uncertain and Evolving Policy:** P16 contended that, though evolving, government policies in Europe are not equipped to deal with generative AI, as well as the new forms of misinformation arising from it. P13 noted that law around generative AI in India is limited by technicalities and intended primarily to stop citizens from being “cheated” by deepfakes. P26 expressed concern about “impinging on the freedom of speech” if generative AI were overapplied for moderating speech, including that produced by AI models. P8 noted uncertainty about the policies of networks like the IFCN, saying “I think you can’t just use that too much in your work, if you want to stay within the framework, which is super important for us.”

**Skillset Scarcity and Disruption:** Participants consistently noted the short supply of generative AI skills. P35 said this scarcity rendered them unaffordable: “people who do know how to do that

are working in organizations where they’re on a much higher wage than anything that a fact-checking or journalism organization could offer.” P3 said finding tech workers in Africa with generative AI skills was difficult and made harder by the headhunting practices of U.S. tech companies, noting “getting the right skill at a level we can actually afford as an NGO has been a [...] major challenge.”

Many participants reported attempts to build generative AI skills internally, in part to prepare for its disruptive impact on fact-checking workflows. P25 described generative AI skills as important both for efficiently dealing with misinformation and for awareness of the misinforming content that generative AI enables. P27 noted difficulty incorporating generative AI due to fears of displacement by the technology, and of changes to the fact-checking profession: “I think the biggest challenge is how to communicate this in a department [...] It won’t replace us, probably, but it will change our work tremendously.” P28 characterized generative AI as a generational issue for local fact-checkers, noting that aversion to new technology hampers the ability to match promulgators of misinformation: “We are not even yet at the point of being comfortable appearing on videos [...] if you [...] confidently face your cell phone and record what is happening, the same way malign actors are, maybe we would have a fighting chance. And I’m not even at the point of talking about AI yet.”

**Western-Centric Design:** P2, P4, and P11 each said that, while generative AI tools perform well in English, their quality remains poor for low-resource languages, particularly local African languages. P13 described the local languages they dealt with as a low priority for companies developing language models. P12 highlighted the need for Africa-centered partnerships to overcome western biases and lack of access to the data for training African models:

The challenge we have right now is, it’s like the new shiny toy and everybody is talking about AI in Africa, but when you actually ask from an African point of view [...], the problem comes with [...] integrated bias in Eurocentric, or let me just say American platforms. So Meta, Google, they’re all developed there and then they’re used here. So the inherent biases in that, that’s not something that we can do anything about [...] it’ll be good to find ways, at least for African tech,

to partner with these people, to develop the tools that would work for the continent. Otherwise, it's just noise.

P28 connects AI language with inauthentic colonial power: “In post-colonial places like ours, we are taught to read in another language. It’s monotonous. And what I’m getting at is that that monotony is the same sound you hear in AI, and it appears very authentic, even if it’s not [...] It really affects the way we are being shaped as a country.”

## 5 DISCUSSION

Realizing opportunities of generative AI in fact-checking requires not only building new technical competencies, but also addressing responsible use in a domain concerned primarily with reliability. We introduce a novel sociotechnical dimension to the design space for generative models that centers information verification; discuss tensions between the values of fact-checking and the values of generative AI; and outline a research agenda for generative AI in fact-checking.

### 5.1 Design Considerations for Generative AI in Fact-Checking

**5.1.1 Designing for Verification.** Every participant centered one concern about generative AI: verification. This arose in Organizational challenges, in reputational risks of publishing unverified output, and in Technological challenges, as fact-checkers must transparently explain their processes and verify model output.

To begin addressing these challenges, we introduce a sociotechnical dimension to the design space of generative models focused on the production and verification of content, conceived using a 2x2 matrix. We locate the *Producer* of content on the X axis, and the *Verifier* of content on the Y axis (see Fig. 4). The upper right (*Human Producer, Human Verifier*) characterizes workflows in most fact-checking and digital media companies, as experienced staff review content authored by junior staff, and editors refine content authored by fact-checkers and writers. The bottom right (*Human Producer, GenAI Verifier*) characterizes the **☒ Quality Assurance** system for catching small errors (P27), and the **☒ Adversarial Analysis** of fact-checking content (P3). It adds security for high-stakes tasks like fact-check publication that are performed primarily by humans. Fact-checkers generally agreed that the bottom left quadrant (*GenAI Producer, GenAI Verifier*), which includes no human oversight, is suitable for low-stakes settings, or where there is a clear evaluation metric that can be used by the verifying model. For example, P2 used generative AI to refine prompts used by other generative models in **► Core NLP Tasks**, improving pipeline performance with little human oversight. Finally, participants held mixed views of the upper left quadrant (*GenAI Producer, Human Verifier*), noting that the **👤 Labor of Output Verification** may render such designs inefficient, or devalue the role of the human. Participants welcomed AI for **► Metadata Generation** and **► Format Conversion** of content into new modalities, noting such uses saved time, even with human review.

Verification is essential for domains where generative AI may provide transformative benefits, but where the consequences of incorrect output are high. Consider applications of generative AI in

law [83], where high-profile mistakes have rendered use of generative models suspect, or medicine, where research suggests demographic biases emerge in models trained on medical text [1]. The verification dimension, envisioned for the high-stakes, information-centered domain of fact-checking, provides a framework for conceptualizing applications that value veracity at least as highly as efficiency.

**5.1.2 Describing Value Tensions.** Value tensions refer to cases wherein the values of a stakeholder in a technology or process come into conflict with the values of another stakeholder or of the technology itself [32]. Drawing on Birhane et al. [4], we do not view generative AI as value-neutral by default, and we conduct a conceptual investigation into the tensions between the values of fact-checking in interaction with generative AI [87]. We use the five principles of the IFCN, an organization founded to promote common standards in fact-checking [60], as a basis for examining the values of fact-checking. Building on insights from our interviews with IFCN signatories and partners, we map each IFCN principle to an underlying value, and describe its tension with generative AI. Table 2 describes these values and value tensions in detail, illustrating conflicts between generative AI and core fact-checking values such as fairness, transparency, and explainability. While many primarily technical tensions are resolvable, especially in the context of human-AI collaboration, tensions surrounding values like Accountability may require significant changes in the relationships between information professionals like fact-checkers and the technology companies that benefit from their labor [43].

**5.1.3 Defining A Research Agenda.** We describe directions for research identified by the participants of our interview study, grouped by the research community (fairness, accountability, or transparency) they primarily address (see Table 3). Participants stressed that researchers at universities and partner organizations will play an important role in advancing these research directions, but that research cannot have meaningful impact without the involvement of fact-checking organizations. P37 noted the limited effort of researchers to connect with fact-checkers, saying, “normally researchers are doing research without even talking with fact-checkers, so they don’t know how the fact-checking world works.” Some directions, such as developing technology for the global south, or combating information inequality, may also require researchers to forge new relationships with individuals and organizations outside of their existing networks.

### 5.2 Limitations and Future Work

Our findings are limited in that we focus solely on fact-checkers, and primarily on IFCN signatories and partners. There are many stakeholders in fact-checking, including audiences for fact-checks, technology providers, government bodies, and indirect beneficiaries of the impact of fact-checking on the information ecosystem [41]. Future work might center the interests and values of these stakeholders. Moreover, while we draw on the literature of organizational change, we are primarily concerned with understanding the evolution of organizations undertaking the work of information verification, rather than organizations broadly. Future work

IFCN Principle	IFCN Description	Value	Tension with Generative AI
Non-Partisanship and Fairness	"Signatory organizations fact-check claims using the same standard for every fact-check. They do not concentrate their fact-checking on any one side. They follow the same process for every fact-check and let the evidence dictate the conclusions. Signatories do not advocate or take policy positions on the issues they fact-check."	Fairness	Generative AI exhibits wide variance depending on the structure of a user's input, and may reflect both implicit and explicit societal biases based on its training and fine-tuning data.
Standards and Transparency of Sources	"Signatories want their readers to be able to verify findings themselves. Signatories provide all sources in enough detail that readers can replicate their work, except in cases where a source's personal security could be compromised. In such cases, signatories provide as much detail as possible."	Transparency	Models cannot affirmatively identify sources, and may hallucinate inaccurate sources of information when asked to do so.
Transparency of Funding and Organization	"Signatory organizations are transparent about their funding sources. If they accept funding from other organizations, they ensure that funders have no influence over the conclusions the fact-checkers reach in their reports. Signatory organizations detail the professional background of all key figures in the organization and explain the organizational structure and legal status. Signatories clearly indicate a way for readers to communicate with them."	Accountability	Models are pretrained using poorly specified data, and may be aligned via practices exploiting low-cost workers in developing countries. Developers deny responsibility to compensate producers of web-scraped data.
Standards and Transparency of Methodology	"Signatories explain the methodology they use to select, research, write, edit, publish and correct their fact-checks. They encourage readers to send claims to fact-check and are transparent on why and how they fact-check."	Explainability	Even when models produce a correct answer, they cannot give a reliable explanation of how it was arrived at.
Open and Honest Corrections Policy	"Signatories publish their corrections policy and follow it scrupulously. They correct clearly and transparently in line with the corrections policy, seeking so far as possible to ensure that readers see the corrected version."	Openness	Generative AI opens a channel to disseminate information not easily observed or corrected by experts.

**Table 2:** We leverage the IFCN principles to identify fact-checking values, and participant insights to describe tensions with generative AI.

<b>Fairness</b>	Directions that seek to mitigate both technical and societal bias and unfairness.	Interest
Technology for the Global South	Building technologies in coordination with fact-checking organizations in the global south to improve model performance and usability, especially in local languages.	P2, 3, 11, 12, 13, 28
Detecting and Mitigating Bias	Developing technical and human-centered approaches to identifying and minimizing bias in model output and human-AI collaborations.	P5, 6, 21, 33, 35
Combating Information Inequality	Developing methods to reach audiences outside of well-educated, well-resourced communities that typically consume fact-checking content.	P4, 9, 17, 29, 30, 31
<b>Accountability</b>	Directions to improve accountability of AI developers to users, fact-checkers to audiences.	Interest
Improving Data Standards and Safety	Developing technical and policy approaches to ensuring that fact-checker data and content is not misused when training or fine-tuning generative AI models.	P27, 29, 31, 37
Auditing for Deceptive Design	Auditing generative AI systems for deceptive design patterns that manipulate human users into placing too much faith in the veracity of their output.	P1, 12, 25, 26, 28
Improving Open Models	Developing highly usable open source and open weight generative AI models to alleviate fact-checker concerns related to the privacy and ownership of their data.	P2, 5, 14, 16, 27, 37
<b>Transparency</b>	Directions to equip fact-checkers with designs and approaches to maximize transparency.	Interest
Benchmark Development	Developing benchmarks that measure language model performance in settings closer to the real-world scenarios faced by fact-checkers, accounting for the novelty of misinformation.	P2, 3, 16, 37, 38
Synthetic Content Detection	Developing approachable generative AI tools for reliably detecting synthetic content, whatever the modality (such as text, image, audio, or video).	P3, 10, 13, 15, 22, 23, 24
Designing for Transparency	Developing design spaces and methodologies that center the transparent processes of fact-checking professionals and organizations.	P27, 31, 36

**Table 3:** We propose nine research directions for generative AI in fact-checking in which our study participants expressed interest.

might seek to generalize or contextualize our findings with other organizations and sectors.

## 6 CONCLUSION

We presented an interview study with  $N=38$  participants at 29 fact-checking organizations across six continents, describing the

opportunities and challenges of incorporating generative AI in sociotechnical fact-checking infrastructures. Insights from interviews formed the basis for a novel Verification dimension in the design space for generative models for fact-checking. The principles of the IFCN informed a description of the value tensions between fact-checking, which centers transparency, fairness, accountability, and reliability, and generative AI, an unpredictable and sometimes

unreliable technology. Finally, we proposed a research agenda for generative technologies, designs, and approaches in fact-checking.

## RESEARCHER POSITIONALITY

The authors are not themselves fact-checkers, but academic researchers who approach the present work with a mix of sociological expertise related to fact-checking work and technical expertise with generative AI. The second author has maintained relationships with the global fact-checking community throughout their career. While we made concerted efforts to faithfully represent the views and experiences of fact-checkers operating globally, including in the global south, we nonetheless recognize that our positionality as academic researchers at a university in a developed country prohibits us from fully capturing the challenges faced and opportunities envisioned by information professionals operating in different contexts from ours.

## ETHICAL CONSIDERATIONS

We note that our focus on IFCN signatories and their partner organizations may not be representative of all fact-checking organizations, including especially disadvantaged organizations with which we would not have interacted via a purposive and snowball sampling methodology. Moreover, we note that while fact-checkers play an essential role in maintaining the health of the information ecosystem in countries with well-established freedom of speech protections, their impact may be limited in less open environments for speech. As a study positioned within the context of a profession that addresses misinforming content, we thus note that we are not able to fully represent the efforts of marginalized individuals and organizations whose work nonetheless improves the health of the information ecosystem; our work extends only to those arenas in which fact-checkers operate.

## ADVERSE IMPACTS

We considered the possibility that our work might elicit more enthusiasm than might be warranted with regard to adopting generative AI in human-centered discipline like fact-checking. While we have devoted large sections of the work to describing challenges, value tensions, and unrealized research, we note explicitly that this work should not be read as an endorsement of generative AI, but as qualitative scientific research describing the perspectives of a community around adopting this new technology. Cautious, human-centered design that centers the values of fact-checkers and the communities they serve will be necessary to ensure that generative AI is not employed for well-meaning but harmful applications in fact-checking, such as rendering the human labor of fact-checking invisible [41] or producing fact-checking content that ultimately lacks the authority of that of a human expert [46], diluting the power of an essential mechanism for maintaining the health of information ecosystems.

## REFERENCES

- [1] Hammaad Adam, Ming Yang, Kenrick D. Cato, Ioana Baldini, Charles R. Senteio, Leo Anthony Celi, Jiaming Zeng, Moninder Singh, and Marzyeh Ghassemi. 2022. Write It Like You See It: Detectable Differences in Clinical Notes by Race Lead to Differential Model Recommendations. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (2022). <https://api.semanticscholar.org/CorpusID:248572183>
- [2] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. , 610–623.
- [3] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Jason Gabriel, and Shakir Mohamed. 2022. Power to the people? opportunities and challenges for participatory AI. *Equity and Access in Algorithms, Mechanisms, and Optimization* (2022), 1–8.
- [4] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The values encoded in machine learning research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. , 173–184.
- [5] Erling Björgvinsson, Pelle Ehn, and Per-Anders Hillgren. 2010. Participatory design and “democratizing innovation”. In *Proceedings of the 11th Biennial Participatory Design Conference* (Sydney, Australia) (*PDC '10*). Association for Computing Machinery, New York, NY, USA, 41–50. <https://doi.org/10.1145/1900441.1900448>
- [6] Virginia Braun and Victoria Clarke. 2022. Everything changes... well some things do: Reflections on, and resources for, reflexive thematic analysis. *QMIP Bulletin* (2022). <https://api.semanticscholar.org/CorpusID:255921405>
- [7] Stephanie Brookes and Lisa Waller. 2023. Communities of practice in the production and resourcing of fact-checking. *Journalism* 24, 9 (2023), 1938–1958.
- [8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [9] Erik Brynjolfsson, Daniel Rock, and Chad Syverson. 2021. The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics* 13, 1 (2021), 333–372.
- [10] Stuart K Card, Jock D Mackinlay, and George G Robertson. 1990. The design space of input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 117–124.
- [11] Taiwan FactCheck Center. 2024. Taiwan FactCheck Center. <https://tfc-taiwan.org.tw/en/>. [Accessed 22-01-2024].
- [12] Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. 2023. Creativity Support in the Age of Large Language Models: An Empirical Study Involving Emerging Writers. *ArXiv abs/2309.12570* (2023). <https://api.semanticscholar.org/CorpusID:262217523>
- [13] Africa Check. 2024. Africa Check. <https://africacheck.org/>. [Accessed 22-01-2024].
- [14] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. 2021. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374* (2021).
- [15] Chequeado. 2024. Chequeado. <https://chequeado.com/>. [Accessed 22-01-2024].
- [16] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The Journal of Positive Psychology* 12 (2017), 297 – 298. <https://api.semanticscholar.org/CorpusID:219624951>
- [17] ColombiaCheck. 2024. ColombiaCheck. <https://colombiacheck.com/>. [Accessed 22-01-2024].
- [18] RMIT FactLab CrossCheck. 2024. RMIT FactLab CrossCheck. <https://www.rmit.edu.au/about/schools-colleges/media-and-communication/industry/factlab/crosscheck>. [Accessed 22-01-2024].
- [19] Anubrata Das, Houjiang Liu, Venelin Kovatchev, and Matthew Lease. 2023. The state of human-centered NLP technology for fact-checking. *Information processing & management* 60, 2 (2023), 103219.
- [20] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (<conf-loc>, <city>Boston-<city>, <state>MA-<state>, <country>USA-<country>, </conf-loc>)* (*EAAMO '23*). Association for Computing Machinery, New York, NY, USA, Article 37, 23 pages. <https://doi.org/10.1145/3617694.3623261>
- [21] Univision El Detector. 2024. Univision El Detector. <https://www.univision.com/especiales/noticias/detector/index.html>. [Accessed 22-01-2024].
- [22] Laurence Dierickx, Carl-Gustav Lindén, and Andreas Lothe Opdahl. 2023. Automated fact-checking to support professional practices: systematic literature review and meta-analysis. *International Journal of Communication* 17 (2023), 21.
- [23] Mats Ekström, Amanda Ramsålv, and Oscar Westlund. 2021. The epistemologies of breaking news. *Journalism Studies* 22, 2 (2021), 174–192.
- [24] Mats Ekström, Amanda Ramsålv, and Oscar Westlund. 2022. Data-driven news work culture: Reconciling tensions in epistemic values and practices of news journalism. *Journalism* 23, 4 (2022), 755–772.
- [25] Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130* (2023).
- [26] Ilker Etikan, Sulaiman Abubakar Musa, Rukayya Sunusi Alkassim, et al. 2016. Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics* 5, 1 (2016), 1–4.
- [27] Factly. 2024. Factly. <https://factly.in/>. [Accessed 22-01-2024].
- [28] Aos Fatos. 2024. Aos Fatos. <https://www.aosfatos.org/>. [Accessed 22-01-2024].

- [29] Science Feedback. 2024. Science Feedback. <https://science.feedback.org/>. [Accessed 22-01-2024].
- [30] Robert G Fichman and Chris F Kemerer. 1999. The illusory diffusion of innovation: An examination of assimilation gaps. *Information systems research* 10, 3 (1999), 255–275.
- [31] Code for Africa. 2024. Code for Africa. <https://github.com/CodeForAfrica/>. [Accessed 22-01-2024].
- [32] Batya Friedman, David G. Hendry, and Alan Borning. 2017. A Survey of Value Sensitive Design Methods. *Found. Trends Hum.-Comput. Interact.* 11, 2 (nov 2017), 63–125. <https://doi.org/10.1561/1100000015>
- [33] Grammarly. 2024. Grammarly. <https://www.grammarly.com/>. [Accessed 22-01-2024].
- [34] Lucas Graves. 2017. Anatomy of a fact check: Objective practice and the contested epistemology of fact checking. *Communication, culture & critique* 10, 3 (2017), 518–537.
- [35] Lucas Graves. 2018. Factsheet: Understanding the promise and limits of automated fact-checking. *Reuters Inst. Study of Journalism, Univ. Oxford, Oxford* (2018).
- [36] Lucas Graves and Michelle A Amazeen. 2019. Fact-checking as idea and practice in journalism. In *Oxford research encyclopedia of communication*.
- [37] Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. 2022. A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics* 10 (2022), 178–206.
- [38] Infoveritas. 2024. Infoveritas. <https://info-veritas.com/>. [Accessed 22-01-2024].
- [39] Shrey Jain, Connor Spelliscy, Samuel Vance-Law, and Scott Moore. 2023. AI and Democracy's Digital Identity Crisis. *arXiv preprint arXiv:2311.16115* (2023).
- [40] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lampe, Lucile Saulnier, et al. 2023. Mistral 7B. *arXiv preprint arXiv:2310.06825*, (2023).
- [41] Prerna Juneja and Tanushree Mitra. 2022. Human and technological infrastructures of fact-checking. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–36.
- [42] S Kapoor and A Narayanan. 2023. How to prepare for the deluge of generative AI on social media.
- [43] Hanlin Li, Nicholas Vincent, Stevie Chancellor, and Brent Hecht. 2023. The Dimensions of Data Labor: A Road Map for Researchers, Activists, and Policy-makers to Empower Data Producers. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1151–1161. <https://doi.org/10.1145/3593013.3594070>
- [44] Litmus. 2024. Litmus. <https://litmus-factcheck.jp/about/en/>. [Accessed 22-01-2024].
- [45] logically.ai. 2024. logically.ai. <https://www.logically.ai/>. [Accessed 22-01-2024].
- [46] Chiara Longoni, Andrey Fradkin, Luca Cian, and Gordon Pennycook. 2022. News from generative artificial intelligence is believed less. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 97–106.
- [47] Maldita. 2024. Maldita. <https://maldita.es/>. [Accessed 22-01-2024].
- [48] Meedan. 2024. Meedan. <https://meedan.com/>. [Accessed 22-01-2024].
- [49] MindNews. 2024. MindNews. <https://www.mindanews.com/>. [Accessed 22-01-2024].
- [50] Meredith Ringel Morris. 2023. Scientists' Perspectives on the Potential for Generative AI in Their Fields. *arXiv preprint arXiv:2304.01420* (2023).
- [51] Meredith Ringel Morris, Carrie J Cai, Jess Holbrook, Chinmay Kulkarni, and Michael Terry. 2023. The design space of generative models. *arXiv preprint arXiv:2304.10547* (2023).
- [52] Mahin Naderifar, Hamideh Goli, and Fereshteh Ghajarie. 2017. Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in development of medical education* 14, 3 (2017).
- [53] Terrence Neumann and Nicholas Wolczynski. 2023. Does AI-Assisted Fact-Checking Disproportionately Benefit Majority Groups Online? In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 480–490.
- [54] Newtral. 2024. Newtral. <https://www.newtral.es/>. [Accessed 22-01-2024].
- [55] OpenAI. 2022. Introducing ChatGPT. *OpenAI Blog*, (Nov 2022), .
- [56] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems* 35 (2022), 27730–27744.
- [57] Angela Phillips. 2010. TRANSPARENCY AND THE NEW ETHICS OF JOURNALISM. *Journalism Practice* 4, 3 (2010), 373–382. <https://doi.org/10.1080/17512781003642972>
- [58] Pagella Politica. 2024. Pagella Politica. <https://pagellapolitica.it/>. [Accessed 22-01-2024].
- [59] Politifact. 2024. Politifact. <https://www.politifact.com/>. [Accessed 22-01-2024].
- [60] Poynter. 2024. International Fact Checking Network. <https://www.poynter.org/ifcn/>. [Accessed 22-01-2024].
- [61] Poynter. 2024. Verified signatories of the IFCN code of principles. <https://ifcncodeofprinciples.poynter.org/signatories>. [Accessed 22-01-2024].
- [62] Kalyan Prasad Agrawal. 2023. Towards adoption of generative AI in organizational settings. *Journal of Computer Information Systems* (2023), 1–16.
- [63] Pravda. 2024. Pravda. <https://pravda.org.pl/>. [Accessed 22-01-2024].
- [64] Australian Associated Press. 2024. Australian Associated Press. <https://www.aap.com.au/>. [Accessed 22-01-2024].
- [65] Agence France Presse. 2024. Agence France Presse. <https://www.afp.com/en>. [Accessed 22-01-2024].
- [66] Organizers Of Queerinai, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claes Voelcker, Danica J. Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, Hetvi J, Huan Zhang, Jaidev Shriram, Kruno Lehman, Luca Soldaini, Maarten Sap, Marc Peter Deisenroth, Maria Leonor Pacheco, Maria Ryskina, Martin Mundt, Milind Agarwal, Nyx Mclean, Pan Xu, A Pranav, Raj Korpan, Ruchira Ray, Sarah Mathew, Sarthal Arora, St John, Tanvi Anand, Vishakha Agrawal, William Agnew, Yanan Long, Zijie J. Wang, Zeerak Talat, Avijit Ghosh, Nathaniel Dennler, Michael Noseworthy, Sharvari Jha, Emi Baylor, Aditya Joshi, Natalia Y. Bilenko, Andrew McNamara, Raphael Gontijo-Lopes, Alex Markham, Evny Dong, Jackie Kay, Manu Saraswat, Nikhil Vytla, and Luke Stark. 2023. Queer In AI: A Case Study in Community-Led Participatory AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1882–1895. <https://doi.org/10.1145/3593013.3594134>
- [67] The Quint. 2024. The Quint. <https://www.thequint.com/international>. [Accessed 22-01-2024].
- [68] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training., (2018), .
- [69] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.
- [70] Evani Radya-Dixit and Gina Neff. 2023. A Sociotechnical Audit: Assessing Police Use of Facial Recognition. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1334–1346.
- [71] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* 1, 2 (2022), 3.
- [72] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*. PMLR, 8821–8831.
- [73] Rappler. 2024. Rappler. <https://www.rappler.com/>. [Accessed 22-01-2024].
- [74] Thomson Reuters. 2024. Thomson Reuters. [https://www.thomsonreuters.com/e\\_n.html](https://www.thomsonreuters.com/e_n.html). [Accessed 22-01-2024].
- [75] Paavo Ritala, Miika Ruokonen, and Laavanya Ramau. 2023. Transforming boundaries: how does ChatGPT change knowledge work? *Journal of Business Strategy* ahead-of-print (2023).
- [76] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.
- [77] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. 2022. Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792* (2022).
- [78] Der Spiegel. 2024. Der Spiegel. <https://www.spiegel.de/international/>. [Accessed 22-01-2024].
- [79] Clay Spinuzzi. 2005. The Methodology of Participatory Design. *Technical Communication* 52 (05 2005), 163–174.
- [80] Lead Stories. 2024. Lead Stories. <https://leadstories.com/>. [Accessed 22-01-2024].
- [81] Charlotte Tang, Yunan Chen, Bryan C Semaan, and Jahmeilah A Roberson. 2015. Restructuring human infrastructure: The impact of EHR deployment in a volunteer-dependent clinic. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 649–661.
- [82] Tech4Peace. 2024. Tech4Peace. <https://t4p.co/>. [Accessed 22-01-2024].
- [83] The New York Times. 2023. The ChatGPT Lawyer Explains Himself. <https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html>. [Accessed 22-01-2024].
- [84] India Today. 2024. India Today. <https://www.indiatoday.in/>. [Accessed 22-01-2024].
- [85] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, . (2023), .
- [86] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, . (2023), .
- [87] Jessica Wolk, Batya Friedman, and Gavin Jancke. 2007. Value Tensions in Design: The Value Sensitive Design, Development, and Appropriation of a Corporation's. *GROUP'07 - Proceedings of the 2007 International ACM Conference on Supporting Group Work*, 281–290. <https://doi.org/10.1145/1316624.1316668>

- [88] Jay Zhangjie Wu, Yixiao Ge, Xintao Wang, Stan Weixian Lei, Yuchao Gu, Yufei Shi, Wynne Hsu, Ying Shan, Xiaohu Qie, and Mike Zheng Shou. 2023. Tune-a-video: One-shot tuning of image diffusion models for text-to-video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7623–7633.
- [89] Frank F Xu, Uri Alon, Graham Neubig, and Vincent Josua Hellendoorn. 2022. A systematic evaluation of large language models of code. In *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming*. 1–10.
- [90] Xinying Yu, Shi Xu, and Mark Ashton. 2023. Antecedents and outcomes of artificial intelligence adoption and application in the workplace: the socio-technical system theory perspective. *Information Technology & People* 36, 1 (2023), 454–474.
- [91] G Zagni and T Canetta. 2023. Generative AI marks the beginning of a new era for disinformation.
- [92] Hubert D Zajac, Dana Li, Xiang Dai, Jonathan F Carlsen, Finn Kensing, and Tariq O Andersen. 2023. Clinician-facing AI in the Wild: Taking Stock of the Sociotechnical Challenges and Opportunities for HCI. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–39.