Probing GPT-4 for Knowledge of Journalistic Tasks

Charlotte Li
Northwestern University
charlotte.li@u.northwestern.edu

Nicholas Diakopoulos
Northwestern University
nad@northwestern.edu

ABSTRACT

"What is an AI system's comprehension of journalism tasks?" This is an important question to ask as conversations around building agents for use in newsrooms are advancing. In order for an AI system to serve as an agent for specific journalism tasks, it must have some understanding of the work of journalism and how tasks within it break down. In this paper, we assess the level of comprehension that GPT-4 has of journalism tasks using work activity and task descriptions from O*NET. We conduct a qualitative analysis of the output from GPT-4 and construct a journalism task taxonomy. We find that the output from GPT-4 covers the majority of descriptions in the baseline and offers new insights into journalistic tasks. We propose recommendations for future practitioner-centric research based on our results.

KEYWORDS

Journalism Work, Task Taxonomy, Agentic AI, GPT-4, Future of Work

ACM Reference Format:
Charlotte Li and Nicholas Diakopoulos. 2024. Probing GPT-4 for Knowledge of Journalistic Tasks. In *Proceedings of Computation + Journalism Symposium (C+J Symposium).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

As AI becomes more integrated in people's everyday life and work, conversations around building agents to assist with tedious tasks in various workspaces, including newsrooms, have also surfaced. So-called "agentic AI" systems have been defined as those that "can pursue complex goals with limited direct supervision. [28]" To be able to autonomously assist with newswork, an AI agent needs to take a high-level directive, like "write me a news article about the latest social trend" and break it down into tasks that it can accomplish towards achieving that larger goal. In other words, it needs to have a sense of the work of journalism and how tasks break down so that it can plan out the work.

The depths of a generative AI model's knowledge of the tasks inherent to producing news should be indicative of its potential to leverage that knowledge as an AI agent. However, it is difficult to test the level of comprehension some commercial state-of-the-art generative AI systems might have for journalism tasks because of the lack of transparency in publicly available training data and the rapid rate at which these systems are being updated. While research in both agent-based frameworks and in the future of work has recently explored the role generative AI can play in different professional settings [13, 17, 34], little has touched on the field of journalistic work, especially when it comes to how generative AI can play a role as a supportive agent.

This paper sets out to address this gap in the current literature by conducting black-box testing [31] on GPT-4's knowledge of tasks that a journalist would do. We use pre-existing work activity descriptions of news analysts, developed by research teams at O*NET [20], as prompt inputs for GPT-4, and we analyzed the tasks outputted by GPT-4 by comparing them against baseline tasks from O*NET and by situating them in related research literature. We find that although news tasks differ depending on one's role in the newsroom, the type of newsroom, and the subject of their reporting, the news tasks output by GPT-4 were sensible. They cover most of the baseline tasks and detailed activities described by O*NET, and they offer an additional task category: journalism training, which was not mentioned by the baseline.

Our findings contribute to computational journalism research by providing a preliminary assessment of GPT-4's knowledge of journalism tasks, a taxonomy of journalistic tasks, and implications for future research directions. The results of the assessment suggest the potential for developing generative AI-powered agents for planning and breaking down tasks in journalism, although an unaddressed question is the degree to which such agents would be able to effectively *carry out* such planned tasks. The taxonomy produced in this experiment can also serve as a framework for conducting human-centric evaluations of journalistic work. Specifically, we recommend that more work be done in understanding success metrics and AI assistant-viability of these tasks.

2 RELATED WORK

The idea of incorporating computer systems as an agent in corporate workflows has been around since the 1990s in data engineering [6]: frameworks for constructing artificial intelligent agents were being developed to support well defined tasks such as information filtering [29] and decision making [5]. With the modern development of machine learning technologies, agent-based frameworks are now actively used for studying a more diverse set of research questions, ranging from misinformation [22] to migration [19]. Additionally, the introduction of generative artificial intelligence greatly expands the task space for intelligent agents, causing implications of generative technologies for agent-based modeling to be an eminent and active topic of research[28]. This paper extends the line of agent-based framework research by situating a state-of-the-art AI system as an agent in the specific domain of journalism and assessing its capability as an agent tasked with newswork.

C+J Symposium, October 25-27, 2024, Boston, MA

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

Another area of relevant research is related to the implications of generative AI in the future of work. Research in this area explores the usage, evaluation, and impact of generative AI in how work gets done. Specifically, some research has been done on how different occupations, such as entrepreneurs [17] and data scientists [13], are changing their work with the introduction of generative AI. In this paper, we contribute to this line of research by assessing how generative AI can be used by journalists for accomplishing various news-production-related tasks. Moreover, given the personalized and varied nature of work, researchers have deployed several different methodologies to analyze the impact of generative AI in work, such as user-centric participatory studies [34] and large scale quantitative approaches [12]. While some of this research touches on the field of journalistic work, it has not studied the kinds of roles generative AI can play in the day-to-day work of a journalist as a supportive agent. This paper is thus an early exploration of the depth of GPT4's knowledge about journalistic work and what it means to incorporate generative agents in a journalism context.

3 METHODOLOGY

3.1 Establishing a Baseline of Occupational Information

In order to understand the quality of GPT-4's knowledge about journalism tasks, we leveraged the O*NET Resource Center [20] for expert-validated occupational information about journalists. Specifically, we looked at the *News Analysts, Reporters, and Journalists* occupation, which includes a ranked list of work activities that tend to be more "important" to the occupation (N=19 work activities), a list of detailed work activities that are performed across a small to moderate number of occupations within a job family (N=15), and a list of tasks that are specific to this occupation (N=30).

We further examined the methods used by O*NET to develop these lists to understand the meaning and proper interpretation of them. The work activity list was developed in 1997 based on a few job analysis questionnaires [23]. Examples of work activities include "Getting Information," "Interpreting the Meaning of Information for Others," and "Thinking Creativity", and they are assigned an importance score based on the O*NET rating scale of 1-5 of importance collected through surveys with practitioners. The task list was developed by "job analyst observations, job holder and supervisor descriptions elicited in group discussions, and task inventories" [14]. Tasks are kept up-to-date with write-ins from practitioners and teams of researchers [8, 11]. Lastly, detailed work activities were developed as an intermediate between occupation-specific tasks and more generalized work activities to compare occupations within the same industry, and an example of detailed work activities looks like "Interview others for news or entertainment purposes."

In our study, we utilize the list of general work activities to prompt GPT-4 for detailed descriptions of journalism tasks to assess its performance against the more detailed work activities and tasks developed by the experts at O*NET.

3.2 Data Collection with GPT-4

We used the following prompt as input to GPT-4 (accessed as the `gpt-4` model name through OpenAI API on March 29th, 2024) to produce a list of tasks that are related to each of the 19 general work activities.

`'For the following work activity, please give one sentence each for as many tasks as possible that a journalist might do that are related to it. Please avoid providing duplicated tasks, and the level of specificity should be consistent across all tasks. Work Activity: {activity name} — {activity description}'`

A few things are worth noting in this prompt. First, we only prompt once per activity while asking GPT-4 to come up with as many tasks as possible related to that activity. This is because when we set out to prompt GPT-4 multiple times for the same activity, it often repeats some tasks across several different responses, introducing an extra difficult step to de-duplicate the data. Additionally, we explicitly asked GPT-4 to create non-repeating tasks in the same level of specificity, in order to obtain results that are unique and conceptually comparable to each other in terms of level of detail.

3.3 Analysis

With the output from GPT-4, we then manually clustered each task, disregarding the corresponding work activity in its prompt. We iterated on inductively labeling and re-clustering these tasks until some larger themes emerge [4], resulting in a journalistic task taxonomy. We further cross-reference this taxonomy with with the O*NET task list and detailed work activities list in order to elaborate on more qualitative differences in our analysis.

4 RESULTS

We obtained 285 task descriptions in total across the top 19 most important activities listed, each activity receiving 10 to 20 tasks. These task descriptions were organized into 6 high level categories: **gathering information**, **sensemaking**, **editing**, **publication and distribution**, **productivity**, and **journalism training**. Each high level category is further divided into several sub-category codes. We present an overview of each high level category for the purpose of analysis below.

- **Gathering information**: This category of tasks includes tasks that aim to monitor, source, and gather information for reporting purposes. Tasks under this category are organized by the *type of sources* involved in information gathering, the *method* used for information gathering, or ways to *maintain sources*. For example, social media is a type of source, for which "A key task could be monitoring social media channels for breaking news and trending topics" would belong. Other source types mentioned include: representatives/spokespeople, experts, scholarly sources, and other media. Methods of information gathering, on the other hand, can include tasks such as interviewing: "A journalist may conduct in-depth interviews with key individuals related to the story." Other methods related to information gathering included crowdsourcing, field work, information requests, translation, and transcription. Lastly, an example of source maintenance looks like "A journalist might establish relationships with insider sources for exclusive information."
- **Sensemaking**: This category consists of tasks that journalists do to make sense of information or concepts for different

purposes. It includes tasks such as *ideation* (e.g., “A journalist might come up with a fresh angle to approach a widely reported story.”); judging the *newsworthiness* of an item (e.g., “A journalist may assess the newsworthiness of a press release or tip.”); *analyzing contents* of different source types (e.g. text or data and statistics) and domains (e.g. politics, law, finance, culture, technology, health, entertainment, science, and history); conducting *archival research*; or making sense of *multimedia* content. Sensemaking tasks may thus intersect with tasks during many other phases of production such as gathering, editing, and publication.

- **Editing**: This category includes tasks that involve checking and editing work in progress for *readability*, *accuracy*, *legality*, and *other journalistic editorial judgments* related to standards and ethics. An example task for checking readability looks like “They might judge the readability and coherence of their own piece before publishing.” An example of checking for accuracy is “They might have to cross-check data from multiple sources to ensure consistency.” For legality it might be, “They may communicate with a legal team to ensure the information they publish is lawful and ethical" and for standards checking an example is, “Journalists may decide on the most ethical manner to present sensitive information to avoid harm and maintain integrity."
- **Publication and Distribution**: This category includes tasks that aim to publish and deliver content to an audience that maximizes information spread and readership. Tasks in this category represent a wide variety of ways to deliver information to readers, such as *live delivery* (e.g., “They may deliver information to the public through live broadcasts or social media updates.”), or publishing content with online formats (e.g., “A journalist could send newsletters or email updates to subscribers, updating them on recent news or articles”). This category also includes ways journalists may *interact and engage with their audience* including via hosting public discussions or engaging through social media or other digital channels in order to understand the impact of their publication and receive feedback. It is important to note here that different technologies, such as newspaper and broadcast have different relationships with publication and distribution. Finally, we included tasks related to *audience analytics* in this category, such as using "analytics tools to track the performance of their articles online."
- **Productivity**: Tasks in this category have the goal of increasing productivity and ensuring the prompt delivery of the work that journalists do. Tasks here include *organizing workspaces and time* (e.g., “They may maintain a digital calendar of upcoming news-worthy events.”); *managing content* to be published (e.g., “They might utilize an online content management system to upload their articles, along with relevant photos or videos.”); *planning* for work and projects (e.g., “They may establish a plan for conducting background research around the topic of their story.”); and *coordinating* with other news employees (e.g., “A journalist may meet in person with their supervisor to discuss concerns or questions about a current assignment”). This category can also intersect with other categories such as publication and distribution, as planning happens with the goal of accomplishing other tasks in mind.
- **Journalism Training**: This category of tasks is related to personal and skill developments for journalists to improve their expertise as a journalist. This includes training in *writing* (e.g., “They could take part in training or a workshop to master a new style of writing or reporting.”); using *technology* (e.g., “They might learn to use a new software for data visualization to make their articles more engaging.”); or participating in training that is general to the work of journalism (e.g., “They would participate in training and development activities with colleagues for team building”).

5 DISCUSSION

Overall, tasks generated by GPT-4 cover a good portion of the detailed work activities (13 of 15) listed on O*NET. The activities that were not covered by the output were both related to operating equipment: “Operate communications, transmissions, or broadcasting equipment” and “Operate still or video cameras or related equipment.” A surprising observation with respect to the detailed work activities was that among the 285 tasks outputted by GPT-4, *writing* as a task is only directly mentioned once in the context of journalism training. In comparison to the task list on O*NET, the output of GPT-4 again covers the majority of the tasks, except “Transmit news stories or reporting information from remote locations, using equipment such as satellite phones, telephones, fax machines, or modems.” Admittedly, this task is reasonably subsumed by digital conveyance of information via the internet in general now.

For tasks that are covered by both O*NET and GPT-4 output, O*NET tends to include several actions as one task, whereas each task produced by GPT-4 is more specific to a particular goal or action. Moreover, O*NET tasks mention “writing” proportionally more (3 of 30) than the GPT-4 output (1 of 285). While this might be attributed to the language used to prompt GPT-4, which was adapted from O*NET general work activities and doesn’t explicitly mention writing as an activity, it might reflect a gap in GPT-4’s understanding of how journalists go about achieving each task. Indeed, although GPT-4’s coverage of the tasks of what constitutes newswork appears to be comprehensive, an open question that our data cannot address is whether GPT-4 would be able to plan to (successfully) undertake those tasks. Another surprising observation is the mention of journalism training in GPT-4 output without keywords related to “training” or “learning” being relayed to GPT-4 during prompting. This is an addition that is not mentioned in either the task list or the detailed work activities list on O*NET.

Knowing the output of GPT-4 aligns rather well with descriptions on O*NET, we next discuss implications and limitations of the task taxonomy. For each of the high level codes, we reviewed existing literature or frameworks that might be useful for understanding goals and successes related to tasks within it. These frameworks might also serve as a basis for constructing agentic systems to assist with those tasks.

An area of research that overlaps with tasks in the category of *gathering information* is Computational News Discovery (CND), though tool prototypes for gathering information also incorporate sensemaking as an integral part [32]. A framework for CND was proposed based on interviews with journalists [10]. Furthermore, some studies and tools were developed to support content discovery [18, 21], but they are limited to certain types of sources and applied domains such as politics or science. Additional types of sources, such as social media and experts and additional sourcing methods such as field work or public record requests invite further exploration for implementation in an agentic system. Moreover, the task of source maintenance as practiced in journalism [25] is an unexplored dimension of computational systems that could benefit agentic systems as they build networks of sources with relationships they can rely on for information discovery.

Sensemaking has been an active area of research within HCI for the past 25 years [26] and continues to attract attention today [15]. An existing framework that may apply to situating sensemaking as a task for journalistic agents is Yang & Soergel's work [37] on a model for individual sensemaking because of its focus on the iterative and cognitive sensemaking process, which aligns well with processes in journalism, though additional work is called for on team-based sensemaking processes. In journalism, aspects of sensemaking that emerged in our taxonomy have been studied computationally including around newsworthiness [21], creativity support for ideation on things like angles of coverage [24] or studies of data journalists and their approach to analysis and use of statistics [33, 36]. At the same time, studies of sensemaking as an overarching process within journalism have seldomly been explored [32]. Further explorations of sensemaking in terms of journalistic values and newsworthiness could be beneficial for building agentic systems in journalism such that how information is made sense of aligns with typical normative or value-based goals within the field [16].

In terms of *editing*, well known AI-commercial tools, such as Grammarly, exist for grammatical editing, and newsroom guidebook-specific copy-editing plugins have also been developed [9, 30]. Research on fact-checking [2] and legality alignment [3] are both active and can be applied to an agentic system in journalism, albeit with important questions about how and whether an autonomous system can be trustworthy enough to make judgements about truth or the alignment with legal interpretation and precedent. A challenge for operationalizing editing for a journalistic AI agent circles back to sensemaking in journalism: determining whether the use of particular language fits certain journalistic values.

With respect to *production and distribution*, two existing areas of research that might be relevant to related tasks are personalization and user engagement, which would help conceptualize audience engagement and audience analytics for a journalism agent. Multimodal models can also help situate different mediums and formats of information delivery (e.g. video generation, voice generation), and the capacity of AI agents for guiding public discussions [1] (online or perhaps even offline) presents an interesting avenue for future research.

A research area worth mentioning for *productivity* is collaborative work. Within it, the Model of Coordinated Action is a framework for conceptualizing collaborative situations [7]. Given the collaborative nature of newswork, such a framework could be helpful for informing an AI agent in partaking in teamwork. Planning is another area of research within productivity that would benefit from further research in terms of agents' capacity to plan larger arcs of content coverage and organize work over longer time periods.

In regards to *journalism training*, some recent research has experimented with the idea of simulating AI agents as coaches [35], though it is not a concept that has been situated in any particular framework. While research on soft skill development may apply partially to a journalism training agent [27], the development of skills that are specific to journalism warrants more research, and could relate to developing a range of different skills needed to support many of the journalistic work tasks mentioned above.

6 LIMITATIONS AND FUTURE WORK

We do not claim that the presented taxonomy is comprehensive of all tasks undertaken in journalism or in news production more specifically. This work is largely an exercise in establishing some notion that a model such as GPT-4 has a reasonable degree of knowledge of what constitutes newswork, though there may still be gaps in what we are able to uncover.

One of the limitations of this experiment is the lack of diversity in types of roles within newsrooms when surveying GPT-4 for tasks. We only prompted GPT-4 for the work "a journalist" might do, thus the outputs we analyzed largely depended on what it associated with the word "journalist." It is evident that newsrooms are highly collaborative and people serve varying roles in newsrooms. Therefore, the tasks they do will differ depending on the roles they occupy (e.g., a news photographer and an audience engagement editor might do very different tasks). Investigating how the outputs from GPT-4 might differ based on different newsroom roles could reveal discrepancies in its understanding of more specific aspects of newswork. Similarly, adjusting other features such as organization types and sizes may or may not impact what GPT-4 determines to be important tasks as well.

Another limitation in this experiment is the lack of practitioner-centered validation and adaptation of this taxonomy. Though we cross-referenced the output from GPT-4 with the tasks on the O*NET database, it is ultimately up to practitioners to assess whether these descriptions are accurate or comprehensive. This experiment would benefit from different types of user studies such as surveys, or ethnography as a comparison metric for the level of specificity and the span of coverage of the output provided by GPT-4 and the resulting taxonomy. It can also be improved upon by situating it within literature on journalistic work and assessing whether the categorizations of journalistic work aligns with that of journalism researchers.

Lastly, given the observed influence of how the descriptions of work activities in the prompt change the outputs of GPT-4, especially by the verbs used to describe tasks. Changing the wording of the activities in the prompts might change the outputs in ways we have not explored yet and could shed light on different prompting strategies when adapting language models into practice.

Overall, this work begins to establish the range of "understanding" GPT-4 might have for the work of journalism. Nevertheless, there are many directions this current taxonomy can extend to. For one, it could serve as a baseline for further investigations of the usage of AI in journalism tasks, such as though a survey. Specifically, we are interested in exploring the current strategies for and attitudes towards completing these tasks, the compatibility for adopting AI-assistance to each of these tasks, and the criteria for evaluating successful performance on these tasks such that human journalists could confidently delegate such tasks to an agentic AI system.

7 CONCLUSION

In this paper, we present an analysis of GPT-4's knowledge about journalistic tasks and compare it to an existing database of occupation descriptions provided by O*NET. We situate the analysis within the context of developing agentic systems for journalism and provide possible frameworks for the development of such systems for specific categories of journalism tasks. We conclude by pointing out the limitations and future directions of this work.

REFERENCES

[1] Lisa P. Argyle, Christopher A. Bail, Ethan C. Busby, Joshua R. Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. 2023. Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences* 120, 41 (2023), e2311627120. https://doi.org/10.1073/pnas.2311627120

[2] Giannis Bekoulis, Christina Papagiannopoulou, and Nikos Deligiannis. 2021. A Review on Fact Extraction and Verification. *ACM Computing Surveys (CSUR)* 55, 1 (11 2021), 1–35. https://doi.org/10.1145/3485127

[3] Elisa Bertino, Ravi Sandhu, Edgar Weippl, Paolo Guarda, Silvio Ranise, and Hari Siswantoro. 2017. Security Analysis and Legal Compliance Checking for the Design of Privacy-friendly Information Systems. *Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies* (6 2017), 247–254.

[4] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological*, Harris Cooper, Paul M. Camic, Debra L. Long, A. T. Panter, David Rindskopf, and Kenneth J. Sher (Eds.). American Psychological Association, 57–71. https://doi.org/10.1037/13620-004

[5] Tung Bui and Jintae Lee. 1999. An agent-based framework for building decision support systems. *Decision Support Systems* 25, 3 (4 1999), 225–237. https://doi.org/10.1016/S0167-9236(99)00008-1

[6] Christoph. Bussler and Stephan. Jablonski. 1994. Implementing agent coordination for workflow management systems using active database systems. *Proceedings of IEEE International Workshop on Research Issues in Data Engineering: Active Databases Systems* (1 1994), 53–59. https://doi.org/10.1109/RIDE.1994.282853

[7] Dan Cosley, Andrea Forte, Luigina Ciolfi, David McDonald, Charlotte P Lee, and Drew Paine. 2015. From The Matrix to a Model of Coordinated Action (MoCA). *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (2 2015), 179–194. https://doi.org/10.1145/2675133.2675161

[8] Jeffrey A. Dahlke and Dan J. Putka. 2021. *Streamlining the Identification of Emerging Tasks in the O*NET System Using Natural Language Processing (NLP): Technical Summary.* Technical Report. National Center for O*NET Development. https://www.onetcenter.org/dl_files/EmergingTasksNLP.pdf

[9] Laura E. Davis and Vishal Saberwal. 2024. Stylebot. https://stylebot.app/

[10] Nicholas Diakopoulos. 2020. Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journalism. *Digital Journalism* 8, 7 (8 2020), 945–967. https://doi.org/10.1080/21670811.2020.1736946

[11] Erich C. Dierdorff and Jennifer J. Norton. 2011. *Summary of Procedures for O*NET Task Updating and New Task Generation.* Technical Report. National Center for O*NET Development. https://www.onetcenter.org/dl_files/TaskUpdating.pdf

[12] 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* (3 2023). arXiv:2303.10130

[13] Ken Gu, Madeleine Grunde-McLaughlin, Andrew McNutt, Jeffrey Heer, and Tim Althoff. 2024. How Do Data Analysts Respond to AI Assistance? A Wizard-of-Oz Study. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (5 2024), 1–22. https://doi.org/10.1145/3613904.3641891

[14] Michael Hubbard, Rodney McCloy, John Campbell, John Nottingham, Phil Lewis, David Rivkin, and Jonathan Levine. 2000. *Revision of O*NET Data Collection Instruments.* Technical Report. National Center for O*NET Development. https://www.onetcenter.org/dl_files/Data_appnd.pdf

[15] Laura Koesten, Kathleen Gregory, Paul Groth, and Elena Simperl. 2021. Talking datasets – Understanding data sensemaking behaviours. *International Journal of Human-Computer Studies* 146 (2021), 102562.

[16] Tomoko Komatsu, Marisela Gutierrez Lopez, Stephann Makri, Colin Porlezza, Glenda Cooper, Andrew MacFarlane, and Sondess Missaoui. 2020. AI should embody our values: Investigating journalistic values to inform AI technology design. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction.* 1–13. https://doi.org/10.1145/3419249.3420105

[17] Yasmine Kotturi, Angel Anderson, Glenn Ford, Michael Skirpan, and Jeffrey P Bigham. 2024. Deconstructing the Veneer of Simplicity: Co-Designing Introductory Generative AI Workshops with Local Entrepreneurs. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (5 2024), 1–16. https://doi.org/10.1145/3613904.3642191 arXiv:2402.17082

[18] Brian McKernan, Jennifer Stromer-Galley, Ania Korsunska, Sarah E. Bolden, Patrícia Rossini, and Jeff Hemsley. 2023. A Human-Centered Design Approach to Creating Tools to Help Journalists Monitor Digital Political Ads: Insights and Challenges. *Digital Journalism* 11, 3 (3 2023), 411–430. https://doi.org/10.1080/21670811.2022.2064321

[19] Zakaria Mehrab, Logan Stundal, Srinivasan Venkatramanan, Samarth Swarup, Bryan Lewis, Henning S Mortveit, Christopher L Barrett, Abhishek Pandey, Chad R Wells, Alison P Galvani, Burton H Singer, David Leblang, Rita R Colwell, and Madhav V Marathe. 2024. An agent-based framework to study forced migration: A case study of Ukraine. *PNAS Nexus* 3, 3 (3 2024), pgae080. https://doi.org/10.1093/pnasnexus/pgae080

[20] National Center for O*NET Development. 2024. O*NET OnLine. Retrieved May 30, 2024, from https://www.onetonline.org/.

[21] Sachita Nishal and Nicholas Diakopoulos. 2022. From Crowd Ratings to Predictive Models of Newsworthiness to Support Science Journalism. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (11 2022), 1–28. https://doi.org/10.1145/3555542

[22] Javier Pastor-Galindo, Pantaleone Nespoli, and Jos A. Ruiprez-Valiente. 2024. Large-Language-Model-Powered Agent-Based Framework for Misinformation and Disinformation Research: Opportunities and Open Challenges. *IEEE Security & Privacy* 22, 3 (1 2024), 24–36. https://doi.org/10.1109/MSEC.2024.3380511 arXiv:2310.07545

[23] Norman G. Peterson, Michael D. Mumford, Walter C. Borman, P. Richard Jeanneret, Edwin A. Fleishman, and Kerry Y Levin. 1997. *O*NET Final Technical Report.* Technical Report. American Institutes for Research, Washington, DC. https://files.eric.ed.gov/fulltext/ED455452.pdf

[24] Savvas Petridis. [n. d.]. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. ([n. d.]).

[25] Zvi Reich. 2006. The Process Model of News Initiative. *Journalism Studies* 7, 4 (2006), 497–514. https://doi.org/10.1080/14616700600757928

[26] Daniel M. Russell, Laura Koesten, Aniket Kittur, Nitesh Goyal, and Michael Xieyang Liu. 2024. Sensemaking: What is it today? *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (5 2024), 1–5.

[27] Kassem A Saleh, Fawaz Al Anzi, Sabina Jeschke, Lars Knipping, Nicole Natho, and Olivier Pfeiffer. 2009. Supporting collaboration in professional soft-skill training courses. *Proceedings of the First Kuwait Conference on e-Services and e-Systems* (11 2009), 1–4. https://doi.org/10.1145/1836029.1836044

[28] Yonadav Shavit, Sandhini Agarwal, Miles Brundage, Steven Adler, Cullen O'Keefe, Rosie Campbell, Teddy Lee, Pamela Mishkin, Tyna Eloundou, Alan Hickey, Katarina Slama, Lama Ahmad, Paul McMillan, Alex Beutel, Alexandre Passos, and David G. Robinson. 2023. *Practices for Governing Agentic AI Systems.* Technical Report. OpenAI. https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf

[29] Katia. Sycara, Anandeep. Pannu, Mike. Willamson, Dajun Zeng, and Keith. Decker. 1996. Distributed intelligent agents. *IEEE Expert* 11, 6 (12 1996), 36–46. https://doi.org/10.1109/64.546581

[30] THE ASSOCIATED PRESS. 2024. AP Style-Checking Tools. https://store.stylebooks.com/apstylechecking.html

[31] Atmesh Mahapatra Abhinandan Singal Divyansh Goel Kaizhu Huang Simone Scardapane Indro Spinelli Mufti Mahmud Amir Hussain Vikas Hassija, Vinay Chamola. 2024. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation* 16 (2024), 45–74.

[32] Yixue Wang and Nicholas Diakopoulos. 2021. Journalistic Source Discovery: Supporting The Identification of News Sources in User Generated Content. In *Proc. Human Factors in Computing Systems (CHI).*

[33] Anthony Van Witsen. 2021. How Science Journalists Verify Numbers and Statistics in News Stories: Towards a Theory. *Journalism Practice* (2021), 1–20. https://doi.org/10.1080/17512786.2021.1947152

[34] Allison Woodruff, Renee Shelby, Patrick Gage Kelley, Steven Rousso-Schindler, Jamila Smith-Loud, and Lauren Wilcox. 2024. How Knowledge Workers Think Generative AI Will (Not) Transform Their Industries. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (5 2024), 1–26. arXiv:2310.06778

[35] Diyi Yang, Caleb Ziems, William Held, Omar Shaikh, Michael S. Bernstein, and John Mitchell. 2024. Social Skill Training with Large Language Models. arXiv:2404.04204 [cs.CL]

[36] Rodrigo Zamith. 2019. Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism. *Digital Journalism* 7, 4 (2019), 1–20.

[37] Pengyi Zhang and Dagobert Soergel. 2020. Cognitive mechanisms in sensemaking: A qualitative user study. *Journal of the Association for Information Science and Technology* 71, 2 (2 2020), 158–171. https://doi.org/10.1002/asi.24221