Related papers at https://jpwahle.com/pub and https://gipplab.org/pub

Preprint of the paper:

Wahle, Jan Philip & Ruas, Terry & Mohammad, Saif M. & Meuschke, Norman & Gipp, Bela, 'AI Usage Cards: Responsibly Reporting AI-Generated Content'. 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2023.

Click to download: BibTeX

AI Usage Cards: Responsibly Reporting AI-generated Content

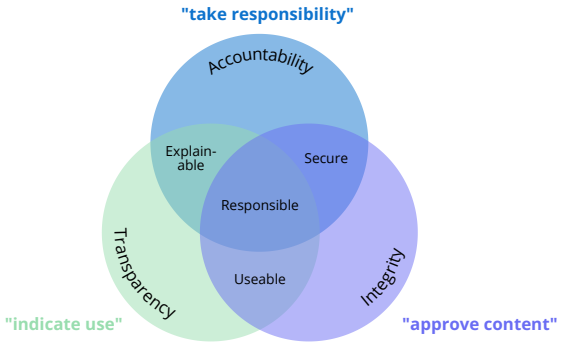
Jan Philip Wahle
University of Göttingen
Göttingen, Germany
wahle@uni-goettingen.de

Terry Ruas
University of Göttingen
Göttingen, Germany
ruas@uni-goettingen.de

Saif M. Mohammad
National Research Council Canada
Ottawa, Canada
saif.mohammad@nrc-cnrc.gc.ca

Norman Meuschke
University of Göttingen
Göttingen, Germany
meuschke@uni-goettingen.de

Bela Gipp
University of Göttingen
Göttingen, Germany
gipp@uni-goettingen.de



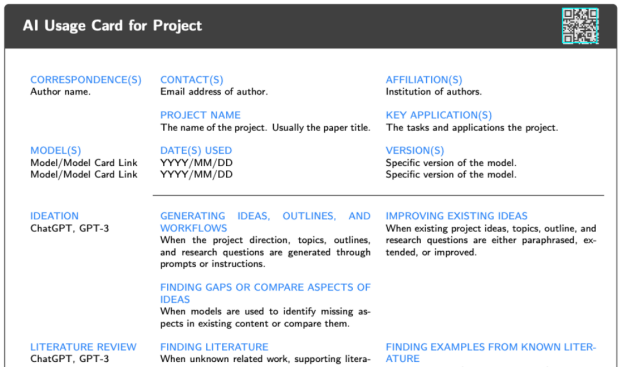


Figure 1: A three-dimensional model of responsible AI usage (left) and the header of the AI Usage Cards template (right).

ABSTRACT

Given AI systems like ChatGPT can generate content that is indistinguishable from human-made work, the responsible use of this technology is a growing concern. Although understanding the benefits and harms of using AI systems requires more time, their rapid and indiscriminate adoption in practice is a reality. Currently, we lack a common framework and language to define and report the responsible use of AI for content generation. Prior work proposed guidelines for using AI in specific scenarios (e.g., robotics or medicine) which are not transferable to conducting and reporting scientific research. Our work makes two contributions: First, we propose a three-dimensional model consisting of transparency, integrity, and accountability to *define* the responsible use of AI. Second, we introduce "AI Usage Cards", a standardized way to *report* the use of AI in scientific research. Our model and cards allow users to reflect on key principles of responsible AI usage. They also help the research community trace, compare, and question various forms of AI usage and support the development of accepted community norms. The proposed framework and reporting system aims to promote the ethical and responsible use of AI in scientific research and provide a standardized approach for reporting AI usage across different research fields. We also provide a free service to easily generate AI Usage Cards for scientific work via a questionnaire and export them in various machine-readable formats for inclusion in different work products at https://ai-cards.org.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; • **General and reference** → **Evaluation**; • **Social and professional topics** → **Computing / technology policy**; • **Software and its engineering** → **Documentation**.

KEYWORDS

ai usage cards, responsible, content generation, text generation, datasheets, model cards, language models, chatgpt

1 INTRODUCTION

The rapid development of AI for natural language processing has led to concerns about using AI-generated content in scientific papers. While AI systems like ChatGPT [8] can generate texts, ideas, and source code, some outputs may be false or fail to attribute the appropriate sources. As a result, it remains unclear whether such models should be used for content generation.

Despite the novelty of AI for content generation, researchers are already using it to support software development and academic writing. However, the rules and norms to govern this new paradigm are yet to be determined. Conferences and journals have started to define rules for using language models to generate content. The largest conference in computational linguistics (ACL) differentiates between AI-based support tools in its official language model policy [9] and calls for *transparency*, i.e., acknowledging the use of

AI when generating content. Conversely, one of the most significant machine learning conferences (ICML) has prohibited using all language models for submissions in 2023 [18]. Although these attempts are not a final and unanimous agreement on using AI-based assistants, they do signal a response to the sudden increase in AI content generation.

While the rules and norms to govern AI-generated content are still emerging, it is clear that transparency is an essential element of responsible use. By acknowledging the use of AI, users can ensure that the generated content is reliable and appropriate attribution is given to the appropriate sources. However, the responsible use of AI for content generation is context-dependent and complex, and caution must be exercised in fields where high-stakes outcomes are involved, such as in the medical and legal fields. Ultimately, while using AI for content generation is likely here to stay, we must find ways to use it responsibly and transparently to avoid adverse outcomes.

History has shown that carefully scoped legalizations can be more effective than blanket prohibitions in many scenarios. Similar to current discussions on social issues such as recreational drug use, a formal framework for acceptable use creates transparency, helps to control the problem, and decriminalizes usage. However, we lack a common language and concrete principles to describe what is needed to "responsibly legalize" AI usage. Although frameworks exist to document key characteristics and ethical considerations associated with models [13], datasets [19], evaluations [6], and AI tasks [14], in the form of cards and sheets, no standardized framework exists to report the use of AI-generated content.

In this work, we introduce the three-dimensional model shown in Figure 1 to decompose the dimensions of responsible AI usage for content generation. We use the term "AI Usage" to refer to all applications of AI-based assistants in scientific work, e.g., to inspire, generate, revise, and compare content. In addition to **transparency** ("Where and how was AI used to produce content?"), our model incorporates **integrity** ("Have humans approved the content produced by AI systems?") and **accountability** ("Who is responsible for the use/dissemination of content produced by AI systems?").

We also propose *AI Usage Cards* as a standard for acknowledging AI support in the scientific process—an important scenario among the many in which content generation can be used. We crafted *AI Usage Cards* with the phases of the scientific process in mind, e.g., hypothesis definition, literature review, and manuscript writing. However, we designed the framework flexible enough to report AI usage across domains. Our cards seek to make the acknowledgment of AI usage the status quo. In summary, this paper makes the following two contributions.

(1) We propose a model for the ***responsible*** use of AI content generation in science based on three dimensions: ***transparency***, ***integrity***, and ***accountability***.
(2) We introduce AI Usage Cards, as a standardized tool to report using AI in scientific works. AI Usage Cards can be generated using an online questionnaire, which we provide as a free service and exports in machine-readable formats that can be incorporated directly into any scientific work.

Our proposed model and method of documenting AI usage can serve as a valuable tool for users to engage in the responsible use of content-generating systems. However, it is important to recognize that potential scenarios for using AI in scientific work can vary significantly, ranging from low-stakes tasks, such as correcting typos in a social media post, to high-stakes scenarios that have the potential to cause harm, such as misleading diagnoses or unjustified sentencing, especially in the medical and legal fields.

In these high-stakes scenarios, AI should be used with extreme caution, and careful consideration should be given to the potential consequences of AI-generated content. There may be applications that we do not want to use AI for at all. In addition, the target audience should also be considered, whether experts or laypeople, as they perceive information as more or less critical. We want to emphasize that simply legalizing AI usage is not the solution; rather, we must find ways to define where and how we should use it and make its usage responsible for avoiding negative consequences.

While people will undoubtedly continue to use AI for content generation, we must take extra care to ensure its responsible use. If used transparently and responsibly, we believe AI can do more good than harm.

2 RELATED WORK

For some time now, studies have shown that humans are increasingly unable to detect AI-generated content [1, 4, 23]. Efforts exist to flag generated text automatically, for example, through zero-shot detection methods [12] or watermarks introduced in generated texts [10]. However, detection techniques alone do not solve important aspects of AI usage, e.g., integrity and accountability.

Concerns regarding the use of AI are also present in the health domain for medical diagnosis [15], clinical trials [2], and health care research [7]. When providing diagnosis, recommending treatment, and screening patients, AI models (such as ChatGPT) must have their predictive capabilities strictly constrained to avoid serious harm, as these models contain biases and unverified information [11]. Moreover, clear accountability for actions taken based on recommendations from AI is required, i.e., "Who is responsible for the diagnosis and treatment?". Although relevant, the contributions of Cruz Rivera et al. [2], Ibrahim et al. [7], Liu et al. [11], Neri et al. [15] are either not transferable to other domains, lack a common agreement, or miss artifacts relevant to our use case in their reports, e.g., tracking which models are used and transparency on modified textual content. Nonetheless, these works mostly agree on the need for better reporting guidelines and more responsible AI.

AI Usage Cards is in line with contributions such as Data Cards [19], Model Cards [13], and Evaluation Cards [6], which complement our work. Data Cards highlight the most important facts about machine learning datasets. The goal of a Data Card is to inform those involved (e.g., researchers) about the characteristics of manipulated datasets during their lifecycle in specific projects (e.g., content, number of records, supporting funding agency). Model Cards [13] document machine learning models—mainly from computer vision and natural language processing—to provide the necessary details for reproducing the results, e.g., training data, metrics, and model parameters. In addition to the technical details on a specific model, Mitchell et al. [13] advocate for transparency of the performance of considered models. Evaluation Cards [6] allow researchers

to document their experiments and standardize evaluations, thereby enabling the monitoring of trends in NLP research.

3 A THREE-DIMENSIONAL MODEL

To responsibly use the outputs of AI, we propose categorizing its usage in a three-dimensional model. The model includes pillars to help acknowledge the usage of AI models (*transparency*); to approve the generated content (*integrity*); and clarify who is responsible for the content generated by the AI model (*accountability*).

While specific frameworks for the use of AI exist in some areas (e.g., public decision-making [3], robotics [22]), they are still limited as they do not specify to which aspects of the scientific work AI systems contributed. Our model is general enough to be applied to many different research domains and still categorizes the main components of any scientific work. In the following, we detail each dimension of our conceptual model and explain how they interact.

Transparency. The first dimension of responsible practice is the foundation on which all others build. Transparency refers to acknowledging the use of AI models in one's work. Throughout the project, one should document every instance of using AI models, with the intent to disclose the information in a suitably summarized form eventually. Similar practices are established for acknowledging a funding agency for supporting a project, individuals for giving feedback, an author affiliation, or a possible conflict of interest. The primary goal of disclosing relevant information about AI usage during scientific work is to establish traceability of the scientific process steps to which AI contributed and help authors reflect on which parts of their work have originated from AI.

Integrity. The second dimension pertains to the correctness, truthfulness, fairness, safety, and appropriateness of the content generated by AI systems. AI uses learnable parameters to model massive amounts of data from the internet. Thus, users of this technology should reflect on such content and assert whether it propagates biases, factual incorrectness, other's ideas as own ideas, unethical statements, prejudice, personally identifiable information, or false claims towards individuals in their output. This verification of fairness principles has to be carried out through human assessment. For smaller initiatives, we recommend obtaining input from a diverse set of people, with representation from relevant groups of people most impacted by system bias. For larger projects, we recommend adopting strategies from participatory design. Participatory design in research and systems development centers the people, especially marginalized and disadvantaged communities, such that they are not merely passive subjects but have the agency to shape the assessment process [5, 16, 17, 21]. All aspects of integrity assessment must be documented for eventual reporting (as appropriate).

Accountability. The last dimension addresses how researchers use, publish, and disseminate content produced with the help of AI models. With clear accountability, those affected by serious harm can ask authors to fix overlooked integrity failures (e.g., propagating hate speech content), especially for those already marginalized. On one side, accountability means ensuring that generated content adheres to integrity proactively before releasing AI-generated content. On the other side, it means being open to feedback and willing to make changes as necessary, and being open about the model's limitations and potential risks associated with its use. Many criticize that AI systems such as ChatGPT are largely black boxes that need more transparency, not least because there has been no official research paper associated with its release. However, transparency is often confused with accountability. When humans perform inexplainable decisions, we can hold them accountable, but if a search engine provides racist results, we cannot. Without accountability, knowing what was manipulated without accountability can produce harm to individuals and does not lead to liability. Therefore, responsible AI use requires documenting corresponding individuals for the project who can be contacted about potential issues.

Responsibility. Ideally, all three dimensions, i.e., transparency, integrity, and accountability, must be fulfilled simultaneously for using AI models responsibly. Analogous to a three-legged chair, each leg/dimension is crucial for using AI in the scientific process. By acknowledging the use of models and holding authors accountable, the output can be explained, but it may still contain inaccuracies and biases since no verification steps have been taken. It is the responsibility of both the AI user and creator to provide sources for the ideas presented, allowing readers to verify them. If we acknowledge the usage of AI and approve its content, we achieve usability but can encounter a lack of accountability if problems occur, e.g., plagiarized content or misleading information. The content is useable "at your own risk" due to a lack of accountability. Finally, when accountability and integrity are verified, the output is secure because it has been affirmed. Even if human flaws exist, they can be accounted for and adjusted if necessary. Only when we acknowledge the use of the model, mitigate its inappropriate outputs, and assign someone to respond in case of problems or questions can we mitigate major issues and make AI usage responsible.

4 AI USAGE CARDS

To facilitate the documentation and reporting needs for the responsible use of AI-generated content, we propose AI Usage Cards as a standardized practice to incorporate the principles of transparency, integrity, and accountability.

The AI Usage Cards presented in this paper are motivated towards using AI in scientific works; however, these cards can easily be adapted to other domains. AI Usage Cards for scientific works follow the scientific method and reflect common headings of scientific manuscripts. We plan to offer different card templates for different disciplines—from computer science to zoology—and stakeholders. Companies, funding agencies, or institutions could define individual questions with varying levels of detail depending on what they deem necessary for ensuring responsible AI usage in their scenario.

AI Usage Cards are available in multiple machine-readable formats, such as LaTeX, XML, JSON, and CSV, to enable easy inclusion in different work products (cf. Figure 1) and support the automated analysis of AI usage across domains and content types in the future. AI Usage Cards can be redistributed for non-commercial purposes according to the CC BY-NC 4.0 license[1]. A free service to create AI Usage Cards[2] is available at

https://ai-cards.org

[1] https://creativecommons.org/licenses/by-nc/4.0/

[2] The AI card reporting usage of AI for this paper can be found on the project webpage.

4.1 Structure

AI Usage Cards are divided into six major blocks: *project details*, *ideation and review*, *methodology and experiments*, *writing and presentation*, *code and data*, and *ethics*. Our card allows researchers and practitioners to report AI usage during all phases of scientific work, from theoretical ideation to practical use. This includes even formulating ideas, for example, in the form of questions and acquiring knowledge during a literature review. Researchers can prompt AI to generate hypotheses, support the development of methods, and suggest newly designed experiments. For measurements, AI can suggest data sources and process data using generated code. Finally, when communicating results, AI can support the writing of papers or other presentation artifacts such as figures and tables. In the following, we summarize each of these blocks:

- **Project details**: Meta-information about the authors, models, and project, including their names, versions, key applications, and affiliations.
- **Ideation and review**: Support during theoretical preparation of the project through idea formulation (e.g., outline) and examining related work (e.g., comparing literature, finding analogies).
- **Methodology and experiments**: Design, comparison, and generation of processes for methodological tasks during the project, optionally resulting in experiments (e.g., proposing a new process).
- **Writing and presentation**: Generation and paraphrasing text used in scientific reports or papers. The improvement of content, figures, tables, and any other elements are also included.
- **Code and data**: Manipulation, generation, refactoring, optimization, and analysis of code and data.
- **Ethics**: Implications of using AI and steps to mitigate their possible errors and harms.

For more details on each category see Appendix C. Each block contains six to seven subcategories. In each subcategory, one or more of the following classifications can be assigned.

- **New:** Content generated based on prompts without a significant portion of prior ideas and thoughts.
  Example Prompt: "Generate some ideas on how to approach the problem of memorization for large language models."
- **Revise:** Content generated based on previous content that has a significant portion of own ideas and thoughts.
  Example Prompt: "Rephrase the following paragraph so that it uses academic voice, is concise and short, and adds an example of a meta-analysis."
- **Compare:** Content generated by providing two or more pieces of content (own or others).
  Example Prompt: "Compare my definition of plagiarism to the following [...]".

We devise a questionnaire to generate AI Usage Cards (see Appendix D for more details). The questions can be answered through a free online form that automatically generates a card to be incorporated in any scientific report as shown in Card 2. The AI card that reports usage for our paper can be found on the project webpage.

5 PRACTICAL CONSIDERATIONS

Q1. Who should create AI Usage Cards?

A. All individuals using any AI system to assist in work are eligible to use AI Usage Cards. The use is intended for a broader audience, from students reporting their assignments at school to researchers submitting research articles and funding proposals.

Q2. When should one create AI Usage Cards?

A. Broadly, AI Usage Cards should be created when using AI systems to support content-related activities during any project (e.g., scientific paper, blog post). This means documenting any activity in which AI has been used as a support tool and categorizing the support according to the card.

Q3. What are the benefits of AI Usage Cards?

A. The benefits of AI Usage Cards include providing transparency, accountability, and integrity in AI usage by encouraging authors to reflect on their AI usage, giving individuals tools to acknowledge AI usage, addressing AI biases for minorities, publishing machine-readable reports for analysis, and contributing to responsible AI usage becoming the status quo.

Q4. What should the cards not be used for?

A. AI Usage Cards should not be used to justify unethical or inappropriate content dissemination. This card should force users to reflect on the generated content and its value to society.

Q5. Should AI Usage Cards be updated?

A. Yes. As technologies change and different domains have varying requirements, the cards must be revised periodically to ensure their relevance and mitigation of outdated practices. For example, the concerns raised by the use of large language models today might not be relevant ten years from now.

Q6. Should we have specific AI Usage Cards?

A. AI Usage Cards should serve as a framework to incorporate different questions for different needs. Given its machine-readable format, AI Usage Cards will allow for data-driven transparency on what was considered important for responsible AI usage by different groups of people throughout time. There might be different cards for different scientific works (e.g., dissertation—more details; short paper—fewer details). Our card might also find applications in other areas, such as philosophy or art, with varying requirements.

Q7. How can we incentivize the creation of AI Usage Cards?

A. Further simplifying the creation of AI Usage Cards, e.g., by offering use-case-specific card formats, can encourage widespread adoption. Moreover, requiring a report on AI usage for any scientific submission, similar to a statutory declaration in graduation theses, can incentivize authors to internalize responsible reporting practices. AI Usage Cards allow creating such reports efficiently in a standardized and machine-processable fashion.

Q8. What role should AI Usage Cards play in the review process of conferences and journals?

A. They should act as documentation to help conference and journal organizers to understand how much support AI systems offer to their users during their work. However, an acceptable threshold for such assistance should be clarified by the organizers.

Q9. Won't the reporting of AI Usage Cards slow the research work?

A. No. AI Usage Cards can be generated in less than five minutes using a free, interactive, and easy-to-use questionnaire. They will help individuals reflect on how they use AI systems in their work. The responsible use of AI and its report is in the interest of all parties involved [14].

6 DISCUSSION

In this paper, we proposed AI Usage Cards to facilitate the reporting of AI use. We provided a free service to produce cards through a fast and interactive questionnaire. We showed how AI use could be reported in a standardized way for all steps of scientific work. Beyond scientific projects, our card serves as a framework to report the usage of AI models across domains.

AI Usage Cards allow to monitor AI usage and help policymakers to evaluate their decisions. Our card can be exported in machine-readable formats (e.g., XML) for further analysis. Compared to other efforts for reporting AI usage [2, 7, 11, 15], we provide a standardized way of reporting that is transferable to other domains.

AI Usage Cards are a tool to acknowledge AI usage openly, similar to how the use of computing infrastructure, external funding, and carbon footprint is acknowledged. The community can decide whether using specific models during certain phases of a scientific project (e.g., hypotheses formulation) is reasonable when certain biases exist. Eventually, analyzing AI Usage Cards across a field of study or area can help to reveal trends in AI usage.

The current version of AI Usage Cards allows for only a limited number of reporters, so we cannot account for many individuals in a scientific project. In this context, focal points or group leaders should take responsibility for reporting their team's actions. The dimensions and properties of AI Usage Cards are a high-level framework for the responsible use and reporting of AI assistance. The questions, subcategories, and technical details of the framework should not remain static in the long run. We invite researchers to build on top of AI Usage Cards and iterate on its concepts. In the same way, scientific contributions are constantly challenged by the community, AI reporting should be regularly updated to guarantee its relevance and mitigation of outdated practices in the community. We leave to future work the investigation on how to adapt AI Usage Cards to the specific needs of its users and projects.

ACKNOWLEDGMENTS

This work was supported by the Lower Saxony Ministry of Science and Culture and the VW Foundation.

REFERENCES

[1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[2] Samantha Cruz Rivera, Xiaoxuan Liu, An-Wen Chan, Alastair K. Denniston, Melanie J. Calvert, The SPIRIT-AI and CONSORT-AI Working Group, SPIRIT-AI and CONSORT-AI Steering Group, Ara Darzi, Christopher Holmes, Christopher Yau, David Moher, Hutan Ashrafian, Jonathan J. Deeks, Lavinia Ferrante di Ruffano, Livia Faes, Pearse A. Keane, Sebastian J. Vollmer, SPIRIT-AI and CONSORT-AI Consensus Group, Aaron Y. Lee, Adrian Jonas, Andre Esteva, Andrew L. Beam, Maria Beatrice Panico, Cecilia S. Lee, Charlotte Haug, Christophe J. Kelly, Christopher Yau, Cynthia Mulrow, Cyrus Espinoza, John Fletcher, David Moher, Dina Paltoo, Elaine Manna, Gary Price, Gary S. Collins, Hugh Harvey, James Matcham, Joao Monteiro, M. Khair ElZarrad, Lavinia Ferrante di Ruffano, Luke Oakden-Rayner, Melissa McCradden, Pearse A. Keane, Richard Savage, Robert Golub, Rupa Sarkar, and Samuel Rowley. 2020. Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension. *Nature Medicine* 26, 9 (Sept. 2020), 1351–1363. https://doi.org/10.1038/s41591-020-1037-7

[3] Karl de Fine Licht and Jenny de Fine Licht. 2020. Artificial Intelligence, Transparency, and Public Decision-Making: Why Explanations Are Key When Trying to Produce Perceived Legitimacy. *AI & SOCIETY* 35, 4 (Dec. 2020), 917–926. https://doi.org/10.1007/s00146-020-00960-w

[4] Biyang Guo, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu. 2023. How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. https://doi.org/10.48550/ARXIV.2301.07597

[5] Beth Humphries, Donna M Mertens, and Carole Truman. 2020. Arguments for an 'emancipatory'research paradigm. In *Research and inequality*. Routledge, 3–23.

[6] Dieuwke Hupkes, Mario Giulianelli, Verna Dankers, Mikel Artetxe, Yanai Elazar, Tiago Pimentel, Christos Christodoulopoulos, Karim Lasri, Naomi Saphra, Arabella Sinclair, Dennis Ulmer, Florian Schottmann, Khuyagbaatar Batsuren, Kaiser Sun, Koustuv Sinha, Leila Khalatbari, Maria Ryskina, Hong Technology, Ryan Cotterell, and Zhijing Jin. 2023. *State-of-the-art generalisation research in NLP: a taxonomy and review*. WorkingPaper. ArXiv. https://doi.org/10.48550/arXiv.2210.03050 We thank Adina Williams, Armand Joulin, Elia Bruni, Lucas Weber, Robert Kirk and Sebastian Riedel for providing us feedback on various stages of this draft, and Gary Marcus for providing detailed feedback on the final draft of this paper. We thank Elte Hupkes for making the app that allows searching through references, and we thank Daniel Haziza and Ece Takmaz for other contributions to the website.

[7] Hussein Ibrahim, Xiaoxuan Liu, and Alastair K. Denniston. 2021. Reporting Guidelines for Artificial Intelligence in Healthcare Research. *Clinical & Experimental Ophthalmology* 49, 5 (July 2021), 470–476. https://doi.org/10.1111/ceo.13943

[8] OpenAI Inc. 2022. ChatGPT: Optimizing Language Models for Dialogue. https://openai.com/blog/chatgpt/. [Online; accessed 24-Jan-2023].

[9] Anna Rogers Jordan Boyd-Graber, Naoaki Okazaki. 2023. ACL 2023 Policy on AI Writing Assistance. https://2023.aclweb.org/blog/ACL-2023-policy/. [Online; accessed 24-Jan-2023].

[10] John Kirchenbauer, Jonas Geiping, Yuxin Wen, Jonathan Katz, Ian Miers, and Tom Goldstein. 2023. A Watermark for Large Language Models. https://doi.org/10.48550/ARXIV.2301.10226

[11] Xiaoxuan Liu, Samantha Cruz Rivera, David Moher, Melanie J. Calvert, Alastair K. Denniston, The SPIRIT-AI and CONSORT-AI Working Group, SPIRIT-AI and CONSORT-AI Steering Group, An-Wen Chan, Ara Darzi, Christopher Holmes, Christopher Yau, Hutan Ashrafian, Jonathan J. Deeks, Lavinia Ferrante di Ruffano, Livia Faes, Pearse A. Keane, Sebastian J. Vollmer, SPIRIT-AI and CONSORT-AI Consensus Group, Aaron Y. Lee, Adrian Jonas, Andre Esteva, Andrew L. Beam, An-Wen Chan, Maria Beatrice Panico, Cecilia S. Lee, Charlotte Haug, Christopher J. Kelly, Christopher Yau, Cynthia Mulrow, Cyrus Espinoza, John Fletcher, Dina Paltoo, Elaine Manna, Gary Price, Gary S. Collins, Hugh Harvey, James Matcham, Joao Monteiro, M. Khair ElZarrad, Lavinia Ferrante di Ruffano, Luke Oakden-Rayner, Melissa McCradden, Pearse A. Keane, Richard Savage, Robert Golub, Rupa Sarkar, and Samuel Rowley. 2020. Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension. *Nature Medicine* 26, 9 (Sept. 2020), 1364–1374. https://doi.org/10.1038/s41591-020-1034-x

[12] Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D. Manning, and Chelsea Finn. 2023. DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. https://doi.org/10.48550/ARXIV.2301.11305

[13] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596

[14] Saif Mohammad. 2022. Ethics Sheets for AI Tasks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*

Papers). Association for Computational Linguistics, Dublin, Ireland, 8368–8379. https://doi.org/10.18653/v1/2022.acl-long.573
[15] Emanuele Neri, Francesca Coppola, Vittorio Miele, Corrado Bibbolino, and Roberto Grassi. 2020. Artificial Intelligence: Who Is Responsible for the Diagnosis? *La radiologia medica* 125, 6 (June 2020), 517–521. https://doi.org/10.1007/s11547-020-01135-9
[16] Lesley-Ann Noel. 2016. Promoting an emancipatory research paradigm in design education and practice. In *Proceedings of DRS2016 International Conference, Vol. 6: Future–Focused Thinking*. Brighton, United Kingdom, 27–30.
[17] Michael Oliver. 1997. Emancipatory research: Realistic goal or impossible dream. *Doing disability research* 2 (1997), 15–31.
[18] International Conference on Machine Learning Program Chairs. 2023. Clarification on Large Language Model Policy LLM. https://icml.cc/Conferences/2023/llm-policy. [Online; accessed 24-Jan-2023].
[19] Mahima Pushkarna, Andrew Zaldivar, and Oddur Kjartansson. 2022. Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1776–1826. https://doi.org/10.1145/3531146.3533231
[20] Timo Spinde, Manuel Plank, Jan-David Krieger, Terry Ruas, Bela Gipp, and Akiko Aizawa. 2021. Neural Media Bias Detection Using Distant Supervision With BABE - Bias Annotations By Experts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. https://doi.org/10.18653/v1/2021.findings-emnlp.101
[21] Clay Spinuzzi. 2005. The methodology of participatory design. *Technical communication* 52, 2 (2005), 163–174.
[22] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Transparent, Explainable, and Accountable AI for Robotics. *Science Robotics* 2, 6 (May 2017), eaan6080. https://doi.org/10.1126/scirobotics.aan6080
[23] Jan Philip Wahle, Terry Ruas, Frederic Kirstein, and Bela Gipp. 2022. How Large Language Models are Transforming Machine-Paraphrase Plagiarism. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 952–963. https://aclanthology.org/2022.emnlp-main.62

A ETHICAL CONSIDERATIONS

As AI language models are trained mostly on human-produced data to generate and manipulate textual content, they are not free of the same issues we are subject to (e.g., discrimination and bias). Thus, using these models might result in inappropriate outcomes that need to be dealt with either during their training phase (e.g., more diverse data) or post-processing (e.g., discarding wrongful text suggestions). The propagation of unethical content (e.g., racial slurs) can lead to destructive consequences, particularly for vulnerable populations [20].

AI Usage Cards cannot guarantee AI will be used to the best of everyone's interest or it will not suggest false claims about sensitive topics and individuals. However, the human component plays a fundamental role in evaluating altered content and ideas. It is our responsibility to "keep in check" the technologies we create and use, so we can take action when necessary. This does not mean our process is error-free. Science is a collective work, so our process requires honesty from all parties involved. Therefore, if researchers using AI models do not report their actions, our card can propagate bias, unethical statements, and prejudice.

B AI USAGE CARDS FOR THIS PAPER

Card 1 reports the AI Usage of this paper. We used AI for suggestions on the name of the cards, for comparing methods on the theoretical model, and to generate a first version of the abstract, which we did not use in the final manuscript.

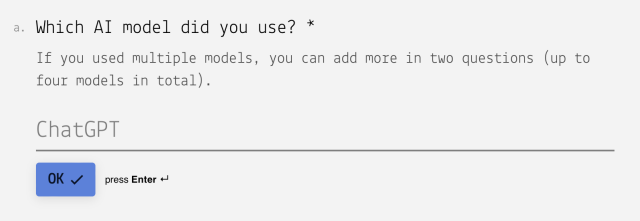


Figure 2: One question of the question set about which AI models were used, when they were used, and in which versions.

C AI USAGE CARDS TEMPLATE

Card 2 shows a template for AI Usage Cards with details on each subcategory. While these categories describe steps of scientific works, they can be adjusted to other domains as well.

D DETAILS ON THE QUESTIONNAIRE

We provide a fast process to generate AI Usage Cards by asking users questions. First, we ask which models have been used in this work by a set of questions, including the model name, date of usage, and model version (see Figure 2 for an example). This set of questions can be repeated multiple times. Next, let users choose the main categories in which AI has been used in their work (see Figure 3). Later, we assign each model to a specific category to differentiate whether one model has been used solely during writing while another one was also used for literature review. The main categories answer the "where" of AI usage. Next, based on the selected main categories, we provide unique views to obtain more details on the "how" of AI usage. For example, has AI been used to generate **new** ideas or to **revise** or **compare**? (Figure 4). For each subcategory, users can provide more details in the form of text in the following steps (Figure 5). They can explain why it was necessary to use AI, how AI has improved the content, which parts of AI content were used, etc. Next, users are presented with ethical questions (Figure 6) and need to approve the used AI-generated content (Figure 7). Finally, a set of questions about the project are presented, including the corresponding authors accountable for the AI usage (see Figure 8 for an example). After submitting the form, the cards are automatically generated and sent by email with a tutorial on how to include them in scientific works (Figure 9).

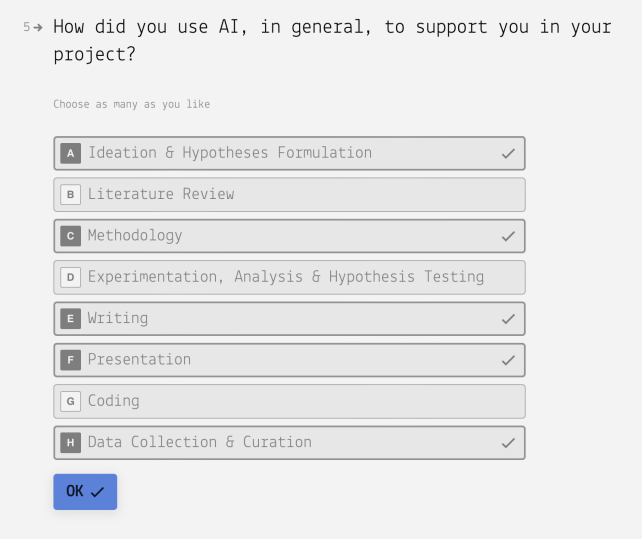


Figure 3: The main categorization in which parts of a work AI was used. This answers the question: "Where was AI used?".

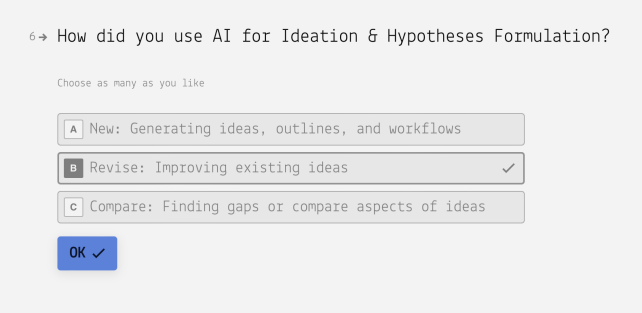


Figure 4: The sub-categories for selected main categories in which AI was used. This answers the question "How was AI used in this specific aspect of the work?"

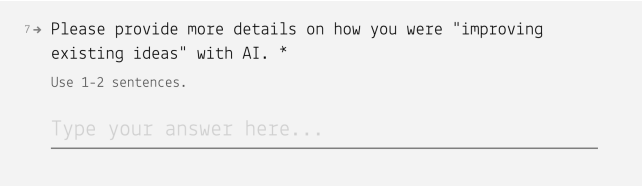


Figure 5: Open text details on each subcategory to mention specifics about how a model was used.

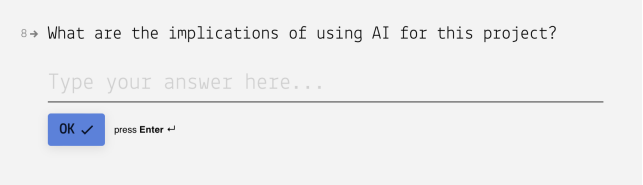


Figure 6: One question of the question set about ethical considerstaions

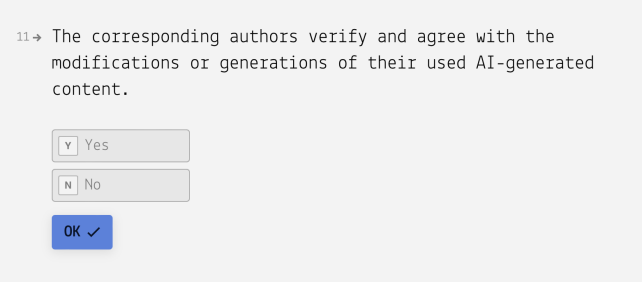


Figure 7: A confirmation that authors approve the used content of AI.

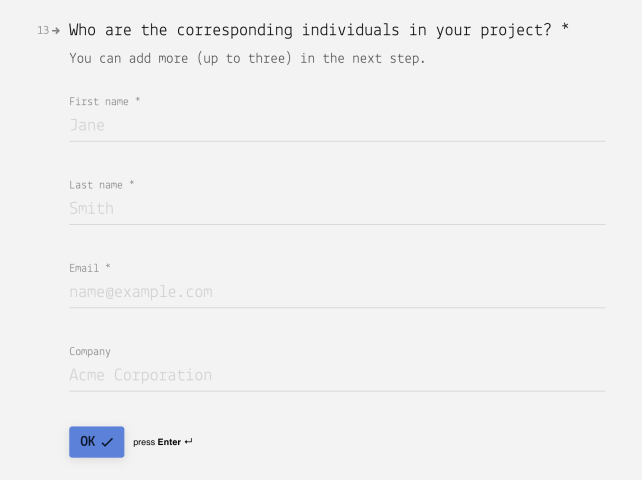


Figure 8: A question of the set for project details to mention the corresponding authors.

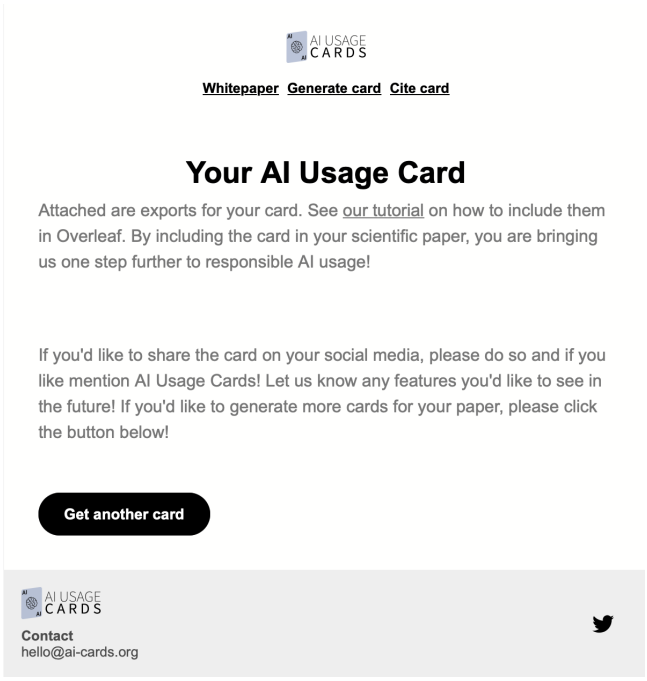


Figure 9: The automatically generated response email to the questionnaire with exports of AI Usage Cards attached and a tutorial on how to include them in scientific works.

AI Usage Cards: Responsibly Reporting AI-generated Content

CORRESPONDENCE(S)
Redacted for anonymity

CONTACT(S)
Redacted for anonymity

AFFILIATION(S)
Redacted for anonymity

PROJECT NAME
AI Usage Cards for Responsibly Reporting Generated Content

KEY APPLICATION(S)
Artificial Intelligence, Reporting, Responsible AI

MODEL(S)
ChatGPT

DATE(S) USED
2023-01-21

VERSION(S)
Not used

IDEATION
ChatGPT

GENERATING IDEAS, OUTLINES, AND WORKFLOWS
Not used

IMPROVING EXISTING IDEAS
Gathering more ideas for the name of AI Usage Cards.

FINDING GAPS OR COMPARE ASPECTS OF IDEAS
Not used

LITERATURE REVIEW

FINDING LITERATURE
Not used

FINDING EXAMPLES FROM KNOWN LITERATURE
Not used

ADDING ADDITIONAL LITERATURE FOR EXISTING STATEMENTS AND FACTS
Not used

COMPARING LITERATURE
Not used

METHODOLOGY
ChatGPT

PROPOSING NEW SOLUTIONS TO PROBLEMS
Not used

FINDING ITERATIVE OPTIMIZATIONS
Not used

COMPARING RELATED SOLUTIONS
Compare multiple versions of our theoretical model.

EXPERIMENTS

DESIGNING NEW EXPERIMENTS
Not used

EDITING EXISTING EXPERIMENTS
Not used

FINDING, COMPARING, AND AGGREGATING RESULTS
Not used

WRITING
ChatGPT

GENERATING NEW TEXT BASED ON INSTRUCTIONS
Generated a first version of the abstract which was not used in the final manuscript.

ASSISTING IN IMPROVING OWN CONTENT
Not used

PARAPHRASING RELATED WORK
Not used

PUTTING OTHER WORKS IN PERSPECTIVE
Not used

PRESENTATION

GENERATING NEW ARTIFACTS
Not used

IMPROVING THE AESTHETICS OF ARTIFACTS
Not used

FINDING RELATIONS BETWEEN OWN OR RELATED ARTIFACTS
Not used

CODING

GENERATING NEW CODE BASED ON DESCRIPTIONS OR EXISTING CODE
Not used

REFACTORING AND OPTIMIZING EXISTING CODE
Not used

COMPARING ASPECTS OF EXISTING CODE
Not used

DATA

SUGGESTING NEW SOURCES FOR DATA COLLECTION
Not used

CLEANING, NORMALIZING, OR STANDARDIZING DATA
Not used

FINDING RELATIONS BETWEEN DATA AND COLLECTION METHODS
Not used

ETHICS ChatGPT	WHAT ARE THE IMPLICATIONS OF USING AI FOR THIS PROJECT? Facilitate the AI usage in scientific work (reporting).	WHAT STEPS ARE WE TAKING TO MITIGATE ERRORS OF AI FOR THIS PROJECT? Careful evaluation of any generated content from the AI model.
	WHAT STEPS ARE WE TAKING TO MINIMIZE THE CHANCE OF HARM OR INAPPROPRIATE USE OF AI FOR THIS PROJECT? Documentation of suggested content in the scientific document.	THE CORRESPONDING AUTHORS VERIFY AND AGREE WITH THE MODIFICATIONS OR GENERATIONS OF THEIR USED AI-GENERATED CONTENT Yes

AI Usage Card v1.0 https://ai-cards.org PDF | BibTeX | XML | JSON | CSV

Card 1: A template for AI Usage Cards.

AI Usage Card for **Project** - Template

CORRESPONDENCE(S) Author name.	CONTACT(S) Email address of author.	AFFILIATION(S) Institution of authors.
	PROJECT NAME The name of the project. Usually, the paper title.	KEY APPLICATION(S) The tasks and applications the project.
MODEL(S) Model/Model Card Link Model/Model Card Link	DATE(S) USED YYYY/MM/DD YYYY/MM/DD	VERSION(S) Specific version of the model. Specific version of the model.
IDEATION ChatGPT, GPT-3, BERT	GENERATING IDEAS, OUTLINES, AND WORKFLOWS When the project direction, topics, outlines, and research questions are generated through prompts or instructions.	IMPROVING EXISTING IDEAS When existing project ideas, topics, outline, and research questions are either paraphrased, extended, or improved.
	FINDING GAPS OR COMPARE ASPECTS OF IDEAS When models are used to identify missing aspects in existing content or compare them.	
LITERATURE REVIEW ChatGPT, GPT-3	FINDING LITERATURE When unknown related work, supporting literature, or similar is obtained through models.	FINDING EXAMPLES FROM KNOWN LITERATURE When examples from a collection of known literature are specified as relevant.
	ADDING ADDITIONAL LITERATURE FOR EXISTING STATEMENTS AND FACTS When literature material is suggested to support existing content.	COMPARING LITERATURE When suggested or existing material is compared and analyzed by the model.
METHODOLOGY RoBERTa	PROPOSING NEW SOLUTIONS TO PROBLEMS When the method and process for solving the problem are outlined.	FINDING ITERATIVE OPTIMIZATIONS When existing method and process are improved.
	COMPARING RELATED SOLUTIONS When existing or generated methods and processes are compared.	
EXPERIMENTS ChatGPT	DESIGNING NEW EXPERIMENTS When new experiment setups are generated through prompts or instructions.	EDITING EXISTING EXPERIMENTS When existing or generated experimental setup is improved.
	FINDING, COMPARING, AND AGGREGATING RESULTS When unseen pa□erns are suggested using existing or generated results to support analysis.	

WRITING GPT-3	GENERATING NEW TEXT BASED ON INSTRUCTIONS When any text is generated through prompts, questions, or instructions.	ASSISTING IN IMPROVING OWN CONTENT When existing text is paraphrased or improved.
	PARAPHRASING RELATED WORK When related work content is paraphrased.	PUTTING OTHER WORKS IN PERSPECTIVE When related work is challenged or paraphrased towards a different direction from their original content.
PRESENTATION DALL E 2, Stable Diffusion	GENERATING NEW ARTIFACTS When new tables, figures, diagrams, or similar elements are generated through instructions or prompts.	IMPROVING THE AESTHETICS OF ARTIFACTS When the visual aspects of tables, figures, diagrams, or similar elements are improved.
	FINDING RELATIONS BETWEEN OWN OR RELATED ARTIFACTS When the content of tables, figures, diagrams, or similar elements are compared to uncover unseen relations.	
CODING PaLM	GENERATING NEW CODE BASED ON DESCRIPTIONS OR EXISTING CODE When new code is generated based on instructions or prompts.	REFACTORING AND OPTIMIZING EXISTING CODE When existing or generated code is refactored or its performance optimized.
	COMPARING ASPECTS OF EXISTING CODE When existing or generated code is compared to uncover unseen patterns or flaws.	
DATA T5	SUGGESTING NEW SOURCES FOR DATA COLLECTION When datasets, collections, or similar sources are suggested based on instructions or prompts.	CLEANING, NORMALIZING, OR STANDARDIZING DATA When any form of noise is removed or mitigated from existing or suggested data.
	FINDING RELATIONS BETWEEN DATA AND COLLECTION METHODS When models are used to establish any relation between datasets' content and collection methods.	
ETHICS	WHAT ARE THE IMPLICATIONS OF USING AI FOR THIS PROJECT? Explain the implications of using AI in the current work scope and its broader impact.	WHAT STEPS ARE WE TAKING TO MITIGATE ERRORS OF AI FOR THIS PROJECT? Explain which decisions and actions were taken to minimize or eliminate the use of AI in this project.
	WHAT STEPS ARE WE TAKING TO MINIMIZE THE CHANCE OF HARM OR INAPPROPRIATE USE OF AI FOR THIS PROJECT? Explain the decisions and actions taken to minimize any form of harm, misuse, and discrimination of the AI model towards any individuals.	THE CORRESPONDING AUTHORS VERIFY AND AGREE WITH THE MODIFICATIONS OR GENERATIONS OF THEIR USED AI-GENERATED CONTENT Verify that any generated or modified content was approved by the authors involved. This can include facts, statements, ideas, and others.

AI Usage Card v1.0 https://ai-cards.org PDF | BibTeX | XML | JSON | CSV

Card 2: A template for AI Usage Cards.