

Communication

Children of AI: A Protocol for Managing the Born-Digital Ephemera Spawned by Generative AI Language Models

Dirk H. R. Spennemann 

School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, Albury, NSW 2640, Australia; dspennemann@csu.edu.au

Abstract: The recent public release of the generative AI language model ChatGPT has captured the public imagination and has resulted in a rapid uptake and widespread experimentation by the general public and academia alike. The number of academic publications focusing on the capabilities as well as practical and ethical implications of generative AI has been growing exponentially. One of the concerns with this unprecedented growth in scholarship related to generative AI, in particular, ChatGPT, is that, in most cases, the raw data, which is the text of the original ‘conversations,’ have not been made available to the audience of the papers and thus cannot be drawn on to assess the veracity of the arguments made and the conclusions drawn therefrom. This paper provides a protocol for the documentation and archiving of these raw data.

Keywords: audio–visual archiving; ChatGPT; cultural heritage digital ephemera; publications ethics



Citation: Spennemann, D.H.R. Children of AI: A Protocol for Managing the Born-Digital Ephemera Spawned by Generative AI Language Models. *Publications* **2023**, *11*, 45. <https://doi.org/10.3390/publications11030045>

Academic Editors: Antony Bryant, Roberto Montemanni, Min Chen, Paolo Bellavista, Kenji Suzuki and Jeanine Treffers-Daller

Received: 28 July 2023

Revised: 4 September 2023

Accepted: 19 September 2023

Published: 21 September 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent months there has been wide-spread public attention regarding the use of artificial intelligence (AI) in various fields. The public releases of the image generator DALL-E and of the generative AI language model ChatGPT (Chat Generative Pre-trained Transformer) in early 2022 caught the public’s imagination. Since then, free-ranging debate emerged regarding the present and potential future abilities of generative AI, the dangers it may pose and the ethics of its usage. ChatGPT is a type of deep learning model that uses transformer architecture to generate coherent and contextually relevant human-like responses based on the input it receives [1].

Since its initial release in 2018, ChatGPT has undergone several revisions, mainly focusing on its increased capabilities in providing longer segments of coherent text and the contextual answering of questions, including the addition of human preferences and feedback. The release of ChatGPT 2.0 in September 2019 relied on a training data set with 1.5 billion parameters, while ChatGPT 3 (June 2020) was trained (by human trainers) on 175 billion parameters. ChatGPT 3.5 was released to the general public to encourage experimentation [2], with a temporal cut for the addition of its training data in September 2021.

ChatGPT has been shown to be capable of producing poetry [3], short stories and plays [4–6], English essays [7], as well as writing lines of code [8]. A growing number of studies has been looking at the effects of generative AI on education and academia, examining the level of knowledge and capabilities of ChatGPT as reflected in its responses to several fields of academic endeavor, such as agriculture [9], archaeology [10], chemistry [11], computer programming [12], cultural heritage management [13], diabetes education [14], digital forensics [15], medicine [16–23], medical education [24], nursing education [25], and remote sensing [10]. ChatGPT is the typical double-edged sword presented by many new technologies [26]. It is a potential tool to enhance student learning e.g., [27–29], while at the same time has the ability to substantively aid students in assignment writing with the associated potential for academic misconduct e.g., [30–32].

2. The Problem

A primary mandate of academic publishing is ethical academic conduct. To ensure the integrity, transparency, and reproducibility of published research, the Committee on Publication Ethics (COPE) has issued a set of guidelines on the deposition and management of data and datasets that were used to explain and substantiate findings that are reported in academic publications [33]. In brief, this entails the deposition of research data in a common readable format in a curated, public, institutional or governmental data depository; the publication of the data and their collection methodology as a stand-alone data publication; or the appendage of the data as supplementary material to the article hosted on the publisher's servers. This mandate is followed by and large with some disciplines being more compliant than others (e.g., medical research).

At this point in time, academic research into ChatGPT and other similar generative AI language models' abilities and limitations is expanding at a rapid rate across most disciplines. The question arises as to how the research data associated with the publications are being managed. While much of the following discussion focusses on ChatGPT, this also applies to the output of other generative AI language models.

The nature of ChatGPT and similar generative AI tools means that each response to a given task will be different. While responses may be structurally and conceptually similar [34], they are not identical. Thus, it is not possible to recreate an identical or near identical response. Furthermore, all conversations with ChatGPT, for example, are deleted after 30 days to maintain server space [35]. Irrespective of this, all conversations with generative AI language models are virtual artefacts (*sensu* [36]), which will eventually disappear due to server upgrades or data warehouse restructuring. Consequently, the original conversations are equivalent to and should be treated like an experiment's raw data, and thus should be archived. At present, academic papers that have been written in relation to ChatGPT, for example, have taken five different approaches:

1. articles that include the entire conversation in the body of the paper [9–11,21,24,34]
2. articles that quote extracts of the conversation in the body of the paper and provide the full text of the conversation as a supplementary document [13] or an appendix [37]
3. articles that quote extracts of the conversation in the text but do not provide access to the full text [15,22,23,25,29,31,38,39]
4. articles that discuss non-quote specific conversation(s) and do not provide access to the full text [30]; and
5. articles that discuss ChatGPT but do not refer to specific conversations, instead discussing the topic at a more abstract level [26,37].

Setting aside the first group, where the conversation makes up the core of the paper, and the second group where the full text is supplied as an appendix or a supplementary file, the other three groups do not allow readers to understand the full context of the conversation and cannot independently assess the validity of the author's interpretation of the interaction.

3. Towards a Solution

3.1. Functional Considerations

While conversations with generative AI language models such as ChatGPT have great similarity with formal interviews in anthropological, and ethnographical and sociological settings [40], they differ on a key aspect related to real-world interactions. For example, human-to-human interviews are conducted in a linear fashion, with one question following on from, and building on, an answer, while in theory, a human respondent could be asked to re-answer the question, and this normally occurs with transitioning phrases [41] and a concomitant second-guessing by the respondent trying to taking cues from the interviewer as to why the previous answer was insufficient (else it would not have been asked in exactly the same way). ChatGPT, on the other hand, allows the human participant to request the generative AI language model to answer the question again, generating a new response to the question. Thus any archiving of a conversation with ChatGPT needs to archive all

“regenerations” of a question if they were conducted, while the paper needs to identify which instance of regeneration was being used or cited.

ChatGPT as a service is not static, but is both being continually upgraded in terms of functionality and server performance and due to its ability to ‘learn.’ The latter is reinforced through user feedback, which is solicited both when a user tasks ChatGPT to regenerate a response, with ChatGPT delivering the second version (Figure 1), and by OpenAI staff monitoring selected conversations (“conversations that may be reviewed by our AI trainers to improve our systems”).

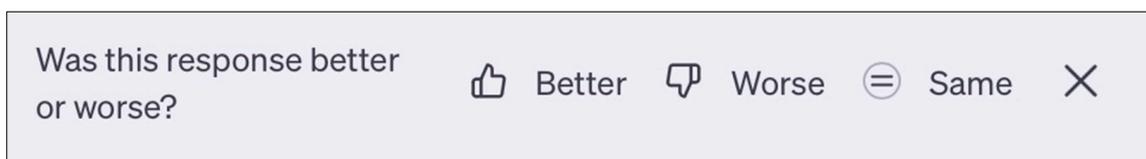


Figure 1. Request for user feedback by ChatGPT following the provision of a regenerated response.

Thus it can be posited that a ChatGPT-like model is time-dependent, and therefore, the date and time of conversation should also be recorded, akin to the practice of stating the access date of web pages in standard referencing.

3.2. Ethical Considerations

Standard ethnographic research practice mandates that interviews are based on a mutual understanding of trust in which the conversation is confidential and any interpretation of the conversation is carried out with expressed and informed consent, even though power dynamics and their changing nature need to be considered [42]. In the current understanding, works created by generative AI do not accrue copyright for the AI system [43] as they fail to meet the human authorship requirement. They can, however, generate copyright for the human actor in the interaction if the latter has substantive guiding involvement [44]. In the same vein, an ‘interview’ with a generative AI language model differs from sociological or ethnographic interviews because generative AI is, at least, at this point in time, not a sentient entity and thus cannot provide informed consent. It follows that, at least, at this point of legal understanding, ‘conversations’ with ChatGPT can be archived.

In traditional ethnographic research practice, notebooks and interview transcripts were commonly deemed personal data, ‘owned’ by the respective researcher. In recent years, research ethics to militate against falsified research findings have led to the mandate to archive and make accessible the original or ‘raw’ research data. In the space of qualitative research, this posed the ethical conundrum to allow access, while, at the same time, maintaining the confidentiality of information provided in the interviews [45–48]. This can be overcome by anonymizing or de-identifying the respondents, although contextual information in the interviews may, in some instances, allow for a re-identification of the informants [49–51].

Conversations with generative AI language models do not fundamentally differ from the transcripts of ethnographic interviews with informants, with the generative AI representing the interviewee. Where privacy issues are involved, for example, where genuine patient records might be used in the assessment of the capabilities of generative AI language models, standard and well-established depersonalization and identity substitution protocols can and should be followed.

It is understood that some research may rely on specific prompt instructions that create specific non-standard outcomes (e.g., prompts that invert the ethical valence [52]), and to what authors may consider ‘proprietary’ for the purpose of ancillary research. In standard scientific research, it is expected that a paper has a methodology section that is public and sets out the research conditions in a way that the experiment can be replicated. Research into and with generative AI language models is no different in this regard.

Consequently, all interactions with generative AI language models that are being analyzed and used in a research paper represent the original research data that need to be

treated in the same way as interview data in qualitative research. Just because the data were created by a generative AI language model instead of a human does not make these data any different, and they need to be curated and managed in the same manner.

It is spurious to argue that this would cause an undue (i.e., time-consuming) burden on the researcher. Data management protocols have been established to improve transparency of research and research findings, and to reduce academic misconduct. Treating ‘interview’ data with generative AI language models with a different standard to interview data with human participants would reopen the door to potential misrepresentation of data and possible academic misconduct.

The required data curation would, at the bare minimum, entail their retention and curation in line with data management policies of the academic institution the author(s) is/are affiliated with and the standards of the academic disciplines they are part of. To do so effectively and comprehensively, a uniform minimum protocol for data collection and documentation is suggested. It is accepted that this protocol is limited to the text-based output of generative AI models.

3.3. Protocol

We propose the following five-step process for ChatGPT (and other generative AI language model) data collection and archiving:

(Step 1) record the metadata, comprised of version and version date, which in the case of ChatGPT can be found at the bottom of the interface (Figure 2), as well as the date and time the conversation occurred, using GMT as the standard.

(Step 2) conduct the conversation as required.

(Step 3) add the end time of the conversation to the metadata entry.

(Step 4) copy the text of the conversation into a text editor or word processor and save the file(s), making sure that all iterations of task generation (if any) are captured and identified as such (e.g., Regeneration 1/3, 2/3, etc.).

(Step 5) generate a complete data document that contains the metadata and text of each conversation.

(Step 6) submit the data document to an approved public or institutional data repository or append it to the publication as a supplementary file.

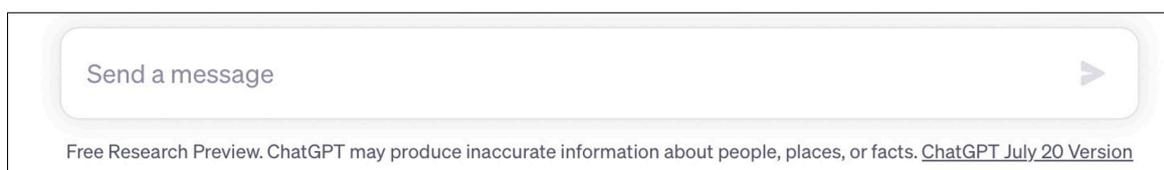


Figure 2. Version date (as shown in the footer of the interaction window).

4. Conclusions

Over the past decade, the development of generative artificial intelligence systems has accelerated dramatically, resulting in the recent public release of the generative AI language model ChatGPT. ChatGPT has captured the public’s imagination with widespread experimentation by academia and the general public alike. Numerous academic disciplines experimented with the capabilities of ChatGPT in relation to their research directions, examining its ability to provide accurate responses. The number of publications on the capabilities of ChatGPT and the practical and ethical implications of the use and abuse of generative AI has been growing exponentially.

This unprecedented growth in scholarship related to generative AI, in particular, ChatGPT, occurs in a large unregulated space, wherein, in most cases, the raw data, which is the text of the original ‘conversations,’ are not being made available to the audience of the papers. In consequence, they cannot be drawn on to assess the veracity of the arguments made in the publications and the conclusions drawn therefrom. This paper has provided a protocol for the documentation and archiving of these raw data.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Markov, T.; Zhang, C.; Agarwal, S.; Eloundou, T.; Lee, T.; Adler, S.; Jiang, A.; Weng, L. New and Improved Content Moderation Tooling. Available online: https://web.archive.org/web/20230130233845mp_/https://openai.com/blog/new-and-improved-content-moderation-tooling/ (accessed on 28 June 2023).
2. Ray, P.P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Phys. Syst.* **2023**, *3*, 121–154. [CrossRef]
3. Moons, P.; Van Bulck, L. ChatGPT: Can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals. *Eur. J. Cardiovasc. Nurs.* **2023**, *2023*, zvad022. [CrossRef]
4. Garrido-Merchán, E.C.; Arroyo-Barrigüete, J.L.; Gozalo-Brihuea, R. Simulating HP Lovecraft horror literature with the ChatGPT large language model. *arXiv* **2023**, arXiv:2305.03429.
5. McGee, R.W. *The Assassination of Hitler and Its Aftermath: A ChatGPT Short Story*; Elsevier: Amsterdam, The Netherlands, 2023; SSRN 4426338.
6. Landa-Blanco, M.; Flores, M.A.; Mercado, M. Human vs. AI Authorship: Does it Matter in Evaluating Creative Writing? A Pilot Study Using ChatGPT. *PsyArXiv* **2023**. [CrossRef]
7. Fitria, T.N. Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *J. Engl. Lang. Teach.* **2023**, *12*, 44–58. [CrossRef]
8. Liu, J.; Xia, C.S.; Wang, Y.; Zhang, L. Is your code generated by chatgpt really correct? rigorous evaluation of large language models for code generation. *arXiv* **2023**, arXiv:2305.01210.
9. Biswas, S. *Importance of Chat GPT in Agriculture: According to Chat GPT*; Elsevier: Amsterdam, The Netherlands, 2023; SSRN 4405391.
10. Agapiou, A.; Lysandrou, V. Interacting with the Artificial Intelligence (AI) Language Model ChatGPT: A Synopsis of Earth Observation and Remote Sensing in Archaeology. *Heritage* **2023**, *6*, 4072–4085. [CrossRef]
11. Castro Nascimento, C.M.; Pimentel, A.S. Do Large Language Models Understand Chemistry? A Conversation with ChatGPT. *J. Chem. Inf. Model.* **2023**, *63*, 1649–1655. [CrossRef]
12. Surameery, N.M.S.; Shakor, M.Y. Use chat GPT to solve programming bugs. *Int. J. Inf. Technol. Comput. Eng. (IJITC)* **2023**, *3*, 17–22, ISSN 2455-5290. [CrossRef]
13. Spennemann, D.H.R. ChatGPT and the generation of digitally born “knowledge”: How does a generative AI language model interpret cultural heritage values? *Knowledge* **2023**, *3*, 480–512. [CrossRef]
14. Sng, G.G.R.; Tung, J.Y.M.; Lim, D.Y.Z.; Bee, Y.M. Potential and pitfalls of ChatGPT and natural-language artificial intelligence models for diabetes education. *Diabetes Care* **2023**, *46*, e103–e105. [CrossRef]
15. Scanlon, M.; Breiting, F.; Hargreaves, C.; Hilgert, J.-N.; Sheppard, J. ChatGPT for Digital Forensic Investigation: The Good, The Bad, and The Unknown. *Preprints* **2023**, 2023070766. [CrossRef]
16. King, M.R. The future of AI in medicine: A perspective from a Chatbot. *Ann. Biomed. Eng.* **2023**, *51*, 291–295. [CrossRef]
17. Sarraju, A.; Bruemmer, D.; Van Iterson, E.; Cho, L.; Rodriguez, F.; Laffin, L. Appropriateness of Cardiovascular Disease Prevention Recommendations Obtained From a Popular Online Chat-Based Artificial Intelligence Model. *JAMA* **2023**, *329*, 842–844. [CrossRef]
18. Bays, H.E.; Fitch, A.; Cuda, S.; Gonsahn-Bollie, S.; Rickey, E.; Hablutz, J.; Coy, R.; Censani, M. Artificial intelligence and obesity management: An Obesity Medicine Association (OMA) Clinical Practice Statement (CPS) 2023. *Obes. Pillars* **2023**, *6*, 100065. [CrossRef]
19. Grünebaum, A.; Chervenak, J.; Pollet, S.L.; Katz, A.; Chervenak, F.A. The exciting potential for ChatGPT in obstetrics and gynecology. *Am. J. Obstet. Gynecol.* **2023**, *228*, 696–705. [CrossRef]
20. Rao, A.S.; Pang, M.; Kim, J.; Kaminen, M.; Lie, W.; Prasad, A.K.; Landman, A.; Dryer, K.; Succi, M.D. Assessing the utility of ChatGPT throughout the entire clinical workflow. *medRxiv* **2023**, *25*, e48659.
21. Sabry Abdel-Messih, M.; Kamel Boulos, M.N. ChatGPT in Clinical Toxicology. *JMIR Med. Educ.* **2023**, *9*, e46876. [CrossRef]
22. Zhu, Y.; Han, D.; Chen, S.; Zeng, F.; Wang, C. How Can ChatGPT Benefit Pharmacy: A Case Report on Review Writing. *Preprints* **2023**, 2023020324. [CrossRef]
23. Haver, H.L.; Ambinder, E.B.; Bahl, M.; Oluyemi, E.T.; Jeudy, J.; Yi, P.H. Appropriateness of Breast Cancer Prevention and Screening Recommendations Provided by ChatGPT. *Radiology* **2023**, *307*, e230424. [CrossRef]
24. Eysenbach, G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. *JMIR Med. Educ.* **2023**, *9*, e46885. [CrossRef] [PubMed]
25. Qi, X.; Zhu, Z.; Wu, B. The promise and peril of ChatGPT in geriatric nursing education: What We know and do not know. *Aging Health Res.* **2023**, *3*, 100136. [CrossRef]

26. Malik, T.; Dwivedi, Y.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Baabdullah, A.M.; Koohang, A.; Raghavan, V. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int. J. Inf. Manag.* **2023**, *71*, 102642.
27. Khan, R.A.; Jawaid, M.; Khan, A.R.; Sajjad, M. ChatGPT-Reshaping medical education and clinical management. *Pak. J. Med. Sci.* **2023**, *39*, 605. [CrossRef]
28. Lim, W.M.; Gunasekara, A.; Pallant, J.L.; Pallant, J.I.; Pechenkina, E. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *Int. J. Manag. Educ.* **2023**, *21*, 100790. [CrossRef]
29. Rudolph, J.; Tan, S.; Tan, S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *J. Appl. Learn. Teach.* **2023**, *6*, 1.
30. Ali, K.; Barhom, N.; Marino, F.T.; Duggal, M. The Thrills and Chills of ChatGPT: Implications for Assessments in Undergraduate Dental Education. *Preprints* **2023**, 2023020513. [CrossRef]
31. Currie, G.; Singh, C.; Nelson, T.; Nabasenja, C.; Al-Hayek, Y.; Spuur, K. ChatGPT in medical imaging higher education. *Radiography* **2023**, *29*, 792–799. [CrossRef]
32. Stokel-Walker, C. AI bot ChatGPT writes smart essays—Should professors worry? *Nature* **2022**. [CrossRef]
33. COPE. Data and Reproducibility. Available online: <https://publicationethics.org/data> (accessed on 12 August 2023).
34. Spennemann, D.H.R. Exhibiting the Heritage of COVID-19—A Conversation with ChatGPT. *Heritage* **2023**, *6*, 5732–5749. [CrossRef]
35. Joshua, J. Data Controls FAQ. Available online: <https://help.openai.com/en/articles/7730893-data-controls-faq> (accessed on 21 July 2023).
36. Spennemann, D.H.R. The Digital Heritage of the battle to contain COVID-19 in Australia and its implications for Heritage Studies. *Heritage* **2023**, *6*, 3864–3884. [CrossRef]
37. Shen, Y.; Heacock, L.; Elias, J.; Hentel, K.D.; Reig, B.; Shih, G.; Moy, L. ChatGPT and Other Large Language Models Are Double-edged Swords. *Radiology* **2023**, *307*, e230163. [CrossRef]
38. Gilson, A.; Safranek, C.W.; Huang, T.; Socrates, V.; Chi, L.; Taylor, R.A.; Chartash, D. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med. Educ.* **2023**, *9*, e45312. [CrossRef]
39. Farhat, F. ChatGPT as a Complementary Mental Health Resource: A Boon or a Bane. *Ann. Biomed. Eng.* **2023**. [CrossRef] [PubMed]
40. Denzin, N.K.; Lincoln, Y.S. *The Sage Handbook of Qualitative Research*; Sage: Newcastle upon Tyne, UK, 2011.
41. Sarantakos, S. *Social Research*, 4th ed.; Macmillan International Higher Education: Basingstoke, UK, 2012.
42. Russell, L.; Barley, R. Ethnography, ethics and ownership of data. *Ethnography* **2020**, *21*, 5–25. [CrossRef]
43. Guadamuz, A. Artificial intelligence and copyright. *Wipo Magazine* **2017**, 14–19. [CrossRef]
44. Copyright Office. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence. Available online: <https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence> (accessed on 23 July 2023).
45. Feldman, S.; Shaw, L. The epistemological and ethical challenges of archiving and sharing qualitative data. *Am. Behav. Sci.* **2019**, *63*, 699–721. [CrossRef]
46. Kuula, A. Methodological and ethical dilemmas of archiving qualitative data. *Assist Q.* **2011**, *34*, 12. [CrossRef]
47. Reeves, J.; Treharne, G.J.; Ratima, M.; Theodore, R.; Edwards, W.; Poulton, R. A one-size-fits-all approach to data-sharing will not suffice in lifecourse research: A grounded theory study of data-sharing from the perspective of participants in a 50-year-old lifecourse study about health and development. *BMC Med. Res. Methodol.* **2023**, *23*, 118. [CrossRef]
48. Richardson, J.C.; Godfrey, B.S. Towards ethical practice in the use of archived transcribed interviews. *Int. J. Soc. Res. Methodol.* **2003**, *6*, 347–355. [CrossRef]
49. Cecaj, A.; Mamei, M.; Bicocchi, N. Re-identification of anonymized CDR datasets using social network data. In Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS), Budapest, Hungary, 24–28 March 2014; pp. 237–242.
50. Bandara, P.K.; Bandara, H.D.; Fernando, S. Evaluation of re-identification risks in data anonymization techniques based on population uniqueness. In Proceedings of the 2020 5th International Conference on Information Technology Research (ICITR), Moratuwa, Sri Lanka, 2–4 December 2020; pp. 1–5.
51. Larbi, I.B.C.; Burchardt, A.; Roller, R. Clinical Text Anonymization, its Influence on Downstream NLP Tasks and the Risk of Re-Identification. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop, Dubrovnik, Croatia, 2–6 May 2023; pp. 105–111.
52. Spennemann, D.H.R. Exploring ethical boundaries: Can ChatGPT be prompted to give advice on how to cheat in university assignments? *Preprint* **2023**, 2023081271. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.