WHO IS RESPONSIBLE FOR INTEGRITY IN THE AGE OF ARTIFICIAL INTELLIGENCE? AN ANALYSIS USING THE EXAMPLE OF ACADEMIC WRITING

Nicolaus Wilder[1], Doris Weßels[1], Johanna Gröpler[1], Andrea Klein[1], Margret Mundorf[1]

[1] *Technical University of Applied Sciences Wildau, Germany*

INTRODUCTION

In this workshop we will focus on the impact of the increasing changes in academic practices due to tools based on Artificial Intelligence (AI). The guiding question for the workshop is who is responsible for ensuring academic integrity in such practices. Our research results so far (Wilder et al., 2021) show that the previous concept of integrity, in which the individual or the collaboration bears ultimate responsibility for both the process and the results of their work, can no longer be sustained in the age of AI. The system of human-machine-cooperation, for example in the production of academic text using AI-based tools, is far too complex for individuals to understand and take responsibility for all the processes involved. This outcome requires a fundamental rethinking of the distribution of responsibility for academic integrity. We have designed an initial model that proposes how responsibility can be distributed across broader shoulders. The aim of the workshop is to present, discuss and further develop this draft and thus to initiate what we believe is an urgently needed discourse on the future of academic integrity in times of AI in the scientific community using the example of AI-based academic writing.

BACKGROUND

The fact that AI will revolutionize traditional academic practices is no longer contentious, in particular, since there have been important technical breakthroughs using AI (The Royal Society, 2017; The Royal Society & The Alan Turing Institute, 2019). However, the pervasive presence of AI even in our lives is often unperceived and there is a lack of conscious awareness for the implications. The quality of AI- based tools for text production based on GPT-3 such as copy.ai oder shortlyai.com has improved significantly, so that the outcomes, for instance translations or texts, often are indistinguishable from human ones (Radford et al., 2019; Scott, 2020) or do even outperform college students (EduRef, 2021). New possibilities open up when one thinks of the active and conscious integration of AI-tools into working processes. In journalism, for instance, AI is already used for research and news production with the objective of both more efficiency and new impulses, but not without a discussion about the ethical questions that comes with the use and yet unpredictable potential of AI-tools (Beckett, 2019). And the use of AI for text production is also becoming increasingly apparent in the context of academic education, as well as the opportunities and risks associated with it (Weßels and Meyer, 2021) such as submitting fully AI-generated texts as examination papers. The first declarations and guidelines for responsible work with AI are being published, however, they are mostly focused on the development of the programs and only give a vague

understanding of how those guidelines should be implemented (Université de Montréal, 2018).

Currently, the responsibility for academic writing lies entirely with the authors, both in terms of the production process and the outcome. This distribution of responsibility must be reconsidered considering that the process from developing to using AI and dealing with the effects of AI-usage involves several actors. In order to avoid the issue of responsibility diffusion in this complex system (DeCamp and Tilburt 2019) by making the human actors involved transparent, we propose the following different responsibility roles for human-machine-collaboration:

1. The "creator" of the AI develops algorithms of a program and models, selects, and provides the set of used data, tests the software, monitors the system etc.
2. The "tool expert" of the AI selects and purposefully configures the AI- application.
3. The "user" of the AI integrates the AI-application to their work as
   - producer who collaborates with the AI consciously.
   - consumer who receives, spreads and comments AI-generated texts.
4. The "affected person" is unaware of the AI's involvement in the process or ignores it and is indirectly affected by the results as a member of the (academic) society.

METHODOLOGY

For the purpose of achieving the specific objectives set for the workshop as efficiently as possible, the workshop will be divided into six steps:

Step 1: Presentation of the status quo The first step is to present the current development of AI in relation to academic text production and to outline the associated challenges for academic integrity. Finally, the first draft of the different roles is presented.

Step 2: Defining the role differentiations After the presentation of the model, it will be discussed with the participants whether the proposed differentiation is appropriate or needs to be adapted.

Step 3: What can each role be responsible for? All workshop participants are assigned to the defined roles in small groups. There they are to determine which specific responsibilities can be covered by this role. The results are recorded.

Step 4: Consolidation of the results in the plenary session Here, each group briefly presents the results they have worked out. All results are presented in key words on a virtual whiteboard.

Step 5: Developing a who-is-where-responsible matrix Ideally, a two-dimensional matrix should be developed in this step which shows at a glance which role is responsible for which aspects of academic integrity.

Step 6: Final discussion Finally, the results of the workshop are summarized and reflected on and the further handling of the results is discussed.

Miroboard will be used for the documentation of the results.

EXPECTED OUTCOMES

The aim of the workshop is to engage in an intensive discourse on the integrity- responsibility model with the participants. At the same time, the draft model is to be subjected to a validity check by the participating experts for academic integrity and further elaborated at the same time. The manifested outcome of the workshop is the further development of the role differentiation into a two-dimensional matrix defining on one axis the different roles responsible for academic integrity in the complex human-machine-collaborations. And on the other axis, the different sub-areas of academic integrity are listed, i.e., what responsibility is taken for. Ideally, the matrix shows at a glance which role is responsible for which aspects. These results will then be presented to the wider community for discussion in a publication in which the participants will be named as collaborators in the development of the matrix.

REFERENCES

BECKETT, CH. (2019). *New powers, new responsibilities. A global survey of journalism and artificial intelligence.* `https://drive.google.com/file/d/1utmAMCmd4rfJHrUfLLfSJ-clpFTjyef1/view`

DECAMP, M., and TILBURT, J. C. (2019). Why we cannot trust artificial intelligence in medicine. *The Lancet Digital Health*, 1(8), E390. `https://doi.org/10.1016/S2589-7500(19)30197-9`

EduRef – Education Reference Desk. (2021). *What Grades Can AI Get In College?.* `https://www.eduref.net/features/what-grades-can-ai-get-in-college/`

RADFORD, A., WU, J., AMODEI, D., CLARK, J., BRUNDAGE, M., and SUTSKEVER, I. (2019). *Better Language Models and Their Implications.* `https://openai.com/blog/better-language-models/`

SCOTT, K. (2020). *Microsoft teams up with OpenAI to exclusively license GPT-3 language model.* `https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/`

The Royal Society (Ed.). (2017). *Machine learning: The power and promise of computers that learn by example.* `https://royalsociety.org/-/media/policy/projects/machine-learning/publications/machine-learning-report.pdf`.

The Royal Society, & The Alan Turing Institute (Eds.). (2019). *The AI revolution in scientific research.* `https://royalsociety.org/-/media/policy/projects/ai-and-society/AI-revolution-in-science.pdf?la=en-GB&hash=5240F21B56364A00053538A0BC29FF5F`.

Université de Montréal (Ed.). (2018). *Montréal Declaration for a Responsible Development of Artificial Intelligence.* `https://5dcfa4bd-f73a-4de5-94d8-c010ee777609.filesusr.com/ugd/ebc3a3_506ea08298cd4f8196635545a16b071d.pdf`.

WESSELS, D., and MEYER, E. (2021). Original oder Plagiat? Der schnelle Weg zur wissenschaftlichen Arbeit im Zeitalter künstlicher Intelligenz. In R. H. REUSSNER, KOZIOLEK A., and HEINRICH R. (Eds.), *Informatik 2020* (pp. 749–761). Bonn: Gesellschaft für Informatik. `https://doi.org/10.18420/inf2020_66`.

WILDER, N., WESSELS, D., GRÖPLER, J., KLEIN, A., and MUNDORF, M. (2021). *Forschungsintegrität und künstliche Intelligenz mit Fokus auf den wissenschaftlichen Schreibprozess. Traditionelle Werte auf dem Prüfstand für eine neue Ära.* Manuscript submitted for publication.