AI Usage Cards: Responsibly Reporting AI-generated Content

Jan Philip Wahle*/**, Terry Ruas*, Saif M. Mohammad**, Norman Meuschke*, Bela Gipp*

*University of Göttingen, Germany – **National Research Council, Canada

Main Takeaways

➢ Disclose OpenAI ChatGPT, Google Bard, etc. in scientific works
➢ Free web service at https://ai-cards.org/ - fill out a quick 5min questionnaire
➢ Exports in machine-readable formats (e.g., LaTeX)
➢ Extendable for different use cases (grant proposals, seminars, assignments, etc.)

Jan Philip Wahle
University of Göttingen

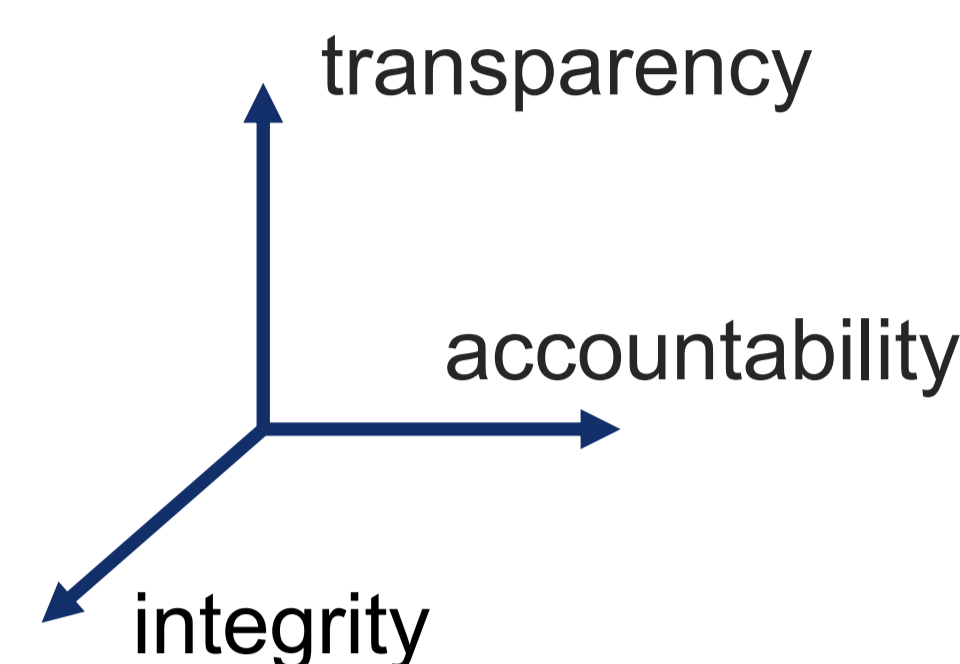✉ wahle@uni-goettingen.de
🐦 @jpwahle

Areas of Interest:
Plagiarism Detection, Paraphrasing

Core Dimensions of AI Usage Cards

- transparency: Where was content generated, and how?
- integrity: How can generations be confirmed?
- accountability: Who takes care of generated content?

Focus on **responsibility**
- core part of the CARE principles[1]
- key element of academic integrity

Existing reporting methods
Model Cards (Mitchell et al., 2018)
Data Sheets (Gebru et al., 2018)
- limited focus on content generation
- not easily applicable to AI-generated content in scientific papers

[1]https://tinyurl.com/rd65u2nr

source: https://academic-englishuk.com/

Components of AI Usage Cards

- based on the scientific method
- categorization using typical paper headings
- coverage of use cases with three primary content operations

Questionnaire

6 → How did you use AI for Ideation & Hypotheses Formulation?

Choose as many as you like

A	New: Generating ideas, outlines, and workflows
B	Revise: Improving existing ideas ✓
C	Compare: Finding gaps or compare aspects of ideas

OK ✓

Content operations

1. **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."
2. **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."
3. **Compare:** Content generated by providing two or more pieces of content (own or others).
   – Example Prompt: "Compare my definition of plagiarism to the following [...]."

AI Usage Card Template

AI Usage Card for *Project* - Template

CORRESPONDENCE(S) CONTACT(S)		AFFILIATION(S)
Author name.	Email address of author.	Institution of authors.
	PROJECT NAME	KEY APPLICATION(S)
	The name of the project. Usually, the paper title.	The tasks and applications the project.
MODEL(S)	DATE(S) USED	VERSION(S)
Model/Model Card Link	YYYY/MM/DD	Specific version of the model.
Model/Model Card Link	YYYY/MM/DD	Specific version of the model.
IDEATION	GENERATING IDEAS, OUTLINES, AND WORKFLOWS	IMPROVING EXISTING IDEAS
ChatGPT, GPT-3, BERT	When the project direction, topics, outlines, and research questions are generated through prompts or instructions.	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.	

Content categories

- **Project details:** Meta-information about authors, models, and targeted project scope.
- **Ideation and review:** Support during the theoretical preparation of the project.
- **Methodology and experiments:** Design, compare, and generate processes for methodological tasks.
- **Writing and presentation:** Generation and paraphrasing text. 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 for using AI and steps to mitigate possible errors and harms.

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

National Research Council Canada   Conseil national de recherches Canada