PhD School on Agriculture, Environment and Bioenergy

(http://sites.unimi.it/dottorato_aab/)

(XL cycle, 2024-27)

Project draft

1. Field of interest

AGR/02 - Agronomia e coltivazioni erbacee

2. Project title

Potential and limits of generative artificial intelligence in agronomic research and good practices for its use.

3. Tutor

Prof. Luca Bechini.

4. Relevance of the topic and state of the art

Large language models (LLMs, like ChatGPT) and other generative artificial intelligence (AI) tools are more and more widespread and powerful. They can generate texts, videos, and images in response to user prompts.

The applications described in the scientific literature, mostly in medical sciences, show that these generative AI tools can support professional tasks (e.g. issue a medical diagnosis; provide nutritional information). In agriculture, a few examples suggest that these tools can support agricultural extension (Tzacor et al., 2023; Ibrahim et al., 2024; Kuska et al., 2024).

It is becoming clear that generative AI can potentially also be used in research, by empowering researchers to carry out time-consuming tasks. For example, Gemini (the LLM developed by Google) could update a literature review on human genetics by scanning 200,000 scientific papers, and extracting and summarizing relevant data from the 250 papers considered relevant (Google, 2023). Another example is that the GPT-3 model was able to predict molecular properties and design molecules based on a target feature (Hocky, 2024). Considering that generative AI can also help with other tasks like preparation of manuscripts and dissemination to wider public of research results through the creation of factsheets and videos, these AI tools have the potential to boost research productivity. This is even more likely considering that their capacity is currently growing. However, so far their potential for agricultural research was not fully explored. Moreover, concerns for inaccurate content generation are raised by many (e.g. Ghassemi et al., 2023).

Therefore, if we want to understand how generative AI can support agricultural researchers in their work, we need to fine-tune the methodology (model training, question framing) and to quantify the accuracy of the results obtained, using rigorous scientific methods. As a results, we will obtain a codified and transparent methodology to obtain optimal results, with the perspective that - as a consequence - researchers will have more time to spend in other tasks not doable by AI.

5 .Layout of the project (draft)

5.1. Materials & Methods

The PhD student will follow an agronomic research project in the field on conservation agriculture. For each research step, she/he will evaluate the capacities of generative AI tools to assist in her/his research work, using a codified and rigorous procedure to train the model, frame the queries, and validate the AI-results against evidence from trusted sources (such as literature data, field results, and process-based simulation models), by measuring the performance of generative AI tools.

The research steps in which the PhD student will test AI tools are:

- Plan the research (literature review): search, read and summarize scientific papers; identify knowledge gaps.
- Process images, e.g. for element counting, pattern identification, color filtering, classification, contour identification.
- Design and carry out statistical analyses; create graphs; develop insights on the results.
- Support the process of manuscript writing by acting as an "intellectual partner" of the researcher.
- Summarize results to identify practical management guidelines and develop decision support rules.
- Disseminate research: e.g. prepare factsheets, presentations and videos for farmers and for the wider public.

The work will be carried out in collaboration with the Department of Computer Science.

5.2. Schedule and major steps (3 years)

- Year 1: literature review; field experiment; data analysis (image processing, statistical analysis).
- Year 2: field experiment; data analysis (image processing, statistical analysis); disseminate research results.
- Year 3: manuscript writing; develop guidelines; disseminate research results.

6. Available funds (to support research)

7PQCP-CSA14LBECH_M, 20,000 € (Luca Bechini)

7. Co-Financing (to support the bourse)

None.

8. Literature

Ghassemi, M., Birhane, A., Bilal, M., Kankaria, S., Malone, C., Mollick, E., Tustumi, F., 2023. ChatGPT one year on: who is using it, how and why? Nature 624, 39–41. doi:10.1038/d41586-023-03798-6

Google, 2023. This is changing the way scientists research | Gemini. https://www.youtube.com/watch?v=sPiOP_CB54A

Hocky, G.M., 2024. Connecting molecular properties with plain language. Nature Machine Intelligence 6, 249–250. doi:10.1038/s42256-024-00812-y

Ibrahim, A., Senthilkumar, K., Saito, K., 2024. Evaluating responses by ChatGPT to farmers' questions on irrigated lowland rice cultivation in Nigeria. Scientific Reports 14, 3407. doi:10.1038/s41598-024-53916-1

Kuska, M.T., Wahabzada, M., Paulus, S., 2024. AI for crop production – Where can large language models (LLMs) provide substantial value? Computers and Electronics in Agriculture 221, 108924. doi:10.1016/j.compag.2024.108924

Tzachor, A., Devare, M., Richards, C., Pypers, P., Ghosh, A., Koo, J., Johal, S., King, B., 2023. Large language models and agricultural extension services. Nature Food 4, 941–948. doi:10.1038/s43016-023-00867-x