Use of Large Language Models and Generative AI Tools in COMAP Contests

This policy is motivated by the rise of large language models (LLMs) and generative AI assisted technologies. The policy aims to provide greater transparency and guidance to teams, advisors, and judges. This policy applies to all aspects of student work, from research and development of models (including code creation) to the written report. Since these emerging technologies are quickly evolving, COMAP will refine this policy as appropriate.

Teams must be open and honest about all their uses of AI tools, including embedded AI (e.g., code-copilots and other AI assisted auto-complete technologies, etc.). The more transparent a team and its submission are, the more likely it is that their work can be fully trusted, appreciated, and correctly used by others. These disclosures aid in understanding the development of intellectual work and in the proper acknowledgement of contributions. Without open and clear citations and references of the role of AI tools, it is more likely that questionable passages and work could be identified as plagiarism and disqualified.

Solving the problems does not require the use of AI tools, although their responsible use is permitted. COMAP recognizes the value of LLMs and generative AI as productivity tools that can help teams in preparing their submission; to generate initial ideas for a structure, for example, or when summarizing, paraphrasing, language polishing etc. There are many tasks in model development where human creativity and teamwork is essential, and where a reliance on AI tools introduces risks. Therefore, we advise caution when using these technologies for tasks such as model selection and building, assisting in the creation of code, interpreting data and results of models, and drawing scientific conclusions.

It is important to note that LLMs and generative AI have limitations and are unable to replace human creativity and critical thinking. COMAP advises teams to be aware of these risks if they choose to use LLMs:

- Objectivity: Previously published content containing racist, sexist, or other biases can arise in LLM-generated text, and some important viewpoints may not be represented.
- Accuracy: LLMs can 'hallucinate' i.e. generate false content, especially when used outside of their domain or when dealing with complex or ambiguous topics. They can generate content that is linguistically but not scientifically plausible, they can get facts wrong, and they have been shown to generate citations that don't exist. Some LLMs are only trained on content published before a particular date and therefore present an incomplete picture.
- Contextual understanding: LLMs cannot apply human understanding to the context of a piece of text, especially when dealing with idiomatic expressions, sarcasm, humor, or metaphorical language. This can lead to errors or misinterpretations in the generated content.
- Training data: LLMs require a large amount of high-quality training data to achieve optimal performance. In some domains or languages, however, such data may not be readily available, thus limiting the usefulness of any output.

Guidance for teams

Teams are required to:

1. **Clearly indicate the use of LLMs or other AI tools, including embedded AI (e.g., code-copilots and other AI assisted auto-complete technologies, etc.) in their report,** including which model was used and for what purpose. Please use inline citations and the reference section. Also append the Report on Use of AI (described below) after your 25-page solution.
2. **Verify the accuracy, validity, and appropriateness** of the content and any citations generated by language models and correct any errors or inconsistencies.
3. **Provide citation and references, following guidance provided here.** Double-check citations to ensure they are accurate and are properly referenced.
4. **Be conscious of the potential for plagiarism** since LLMs may reproduce substantial text from other sources. Check the original sources to be sure you are not plagiarizing someone else's work.

COMAP will take appropriate action
when we identify submissions likely prepared with
undisclosed use of such tools.

Citation and Referencing Directions

Think carefully about how to document and reference whatever tools the team may choose to use. A variety of style guides are beginning to incorporate policies for the citation and referencing of AI tools. Use inline citations and list all AI tools used in the reference section of your 25-page solution.

Whether or not a team chooses to use AI tools, the main solution report is still limited to 25 pages. If a team chooses to utilize AI, following the end of your report, add a new section titled Report on Use of AI. This new section has no page limit and will not be counted as part of the 25-page solution.

Examples (this is *not* exhaustive – adapt these examples to your situation):

Report on Use of AI

1. OpenAI *ChatGPT* (Nov 5, 2023 version, ChatGPT-4)
   Query1: <*insert the exact wording you input into the AI tool*>
   Output: <*insert the complete output from the AI tool*>

2. OpenAI *Ernie* (Nov 5, 2023 version, Ernie 4.0)
   Query1: <*insert the exact wording of any subsequent input into the AI tool*>
   Output: <*insert the complete output from the second query*>

3. Github *CoPilot* (Feb 3, 2024 version)
   Query1: <*insert the exact wording you input into the AI tool*>
   Output: <*insert the complete output from the AI tool*>

4. Google *Bard* (Feb 2, 2024 version)
   Query: <*insert the exact wording of your query*>
   Output: <*insert the complete output from the AI tool*>

5. GitHub *CoPilot* (Jan 16, 2024 version)
   Auto-completions for code used in preparing our models.