Emerging AI tools for Education and Research: Perspective and Policies for IISc

Arindam Khan, Danish Pruthi, Garima Jindal, Kavita Isvaran,
Ramsharan Rangarajan , Siddhartha Gadgil,
Sridharan Devarajan, Suresh Sundaram, Viraj Kumar, Yadati Narahari

I-CARE : IISc Committee on AI Tools for Research and Education - The Mandate

- Explore the challenges and benefits of AI tools in Academic Teaching and Learning
- Provide guidance and recommendations to faculty members on the use of AI tools in teaching and research
- Share best practices from elsewhere
- Explore innovative and ethical uses of AI tools for better learning experiences and outcomes
- Suggest methods for integrity of assessments

Process Followed by the Committee

- Launched a Faculty Survey (received 97 responses) (Oct 2023)
- Launched a Student Survey (received 525 responses) (Oct-Nov 2023)
- Six in-person meetings (Oct-Nov 2023)
- Prepared a 50-page draft document (often comparing our notes with summaries generated by AI tools) (Nov 2023)
- Presentations in the Science and Engineering Faculties and seeking feedback (Today)
- Finalizing the recommendations and submitting the report (Dec. 2023) for immediate implementation after due process

Background on AI Tools

Development of Large Language Models (e.g., ChatGPT)

LLMs are large deep learning models, whose development can be divided into four steps:

1. **Gathering Data & Compute Resources**: models are trained on *trillions of words* using millions of GPU hours (for context, the entirety of Wikipedia is about a billion words)

2. **Pretraining:** models are trained to fill-up the blanked-out words, e.g., "IISc is in the city of ___"

3. **Instruction Tuning:** models are further trained with instructions and (expert-written) outputs. This phase turns these models from next-word predictors to instruction-followers.

4. **Reinforcement Learning from Human Feedback:** models learn from one final layer of feedback provided through human preferences (thumbs up/down on model output).

Capabilities and Limitations of AI Tools

Present Capabilities

- **Writing Assistance:** Can significantly increase productivity; for those with weaker skills, can improve quality.
- **Coding Assistance:** Tools such as Github copilot greatly increase productivity, and especially can easily solve routine programming tasks.
- **Mathematical Reasoning:** Language models show reasonable mathematical reasoning:
  - *GPT-4* scored at the median in the Hungarian national finals math exam (*Claude*, *Grok* were close).
  - At IISc, in *Randomized Algorithms* GPT-4 got 39/60 which was roughly the median score.
  - In *Random Processes,* GPT-4 answered impressively 3 problems which were created by the instructor.
- **Scientific Discovery:** Several recent scientific breakthroughs have harnessed AI tools.

Some Examples of Capabilities (GPT-4)

- **Problem:** A random variable M is uniformly distributed between 1 & 9. For the line with equation y = Mx – $M^{3/2}$ on the x-y plane, what is the probability that it cuts the x-axis after 2?
  Answered correctly.

- **Problem:** Write a Python function to count the number of unique words in a given list of words *of at least length three, ignoring differences between upper-case and lower-case.*
  Answered correctly.

- **Problems** from LeetCode, which is a popular platform for software engineering interview questions
  Answered 68% easy problems, 40% medium problems and 10% hard problems correctly.

- **Problem:** Rewrite the following paragraph. Reduce its length and remove unnecessary adjectives.

  Answered reasonably.

Present Limitations

- AI tools can introduce factual errors, and can hallucinate (convincingly)
- AI tools are poor at judging correctness, limiting their utility for assessment
- Most AI tools are trained on text which is predominantly in English, and
  - Overrepresent concepts, notions and knowledge about certain geographies

- Tools to distinguish AI-generated content from human-written text are *unreliable*
  - But there's hope for the future given recent progress in watermarking models

Guidelines and Best Practices for Research

Recommendations of the Committee

From FAQs of ETH, Zurich

- "Any rapidly written overly specific or restrictive rules and regulations will run the danger of sounding like blind actionism, and apparent quick fixes might sound ridiculous a few months from now."
- "We also need to be careful about restricting tools where we cannot justify it. It makes sense to forbid the use of computers in chess tournaments, since computers could beat humans virtually anytime, which destroys the enjoyment of a wonderful game - this is similar to forbidding doping or "performance-enhancing drugs" in sports. "
- "Science, math and engineering never worked that way; we always worked with the best available tools, and in fact were expected to do so to stay competitive."
- "Having said that, simply copy/pasting the output of an AI-tool and submitting it as one’s own work is unacceptable."

Disclosure and Attribution

- Use of AI tools should be fully disclosed, with attribution that is precise and transparent, and with adequate details.
- When using AI tools in publications, follow the guidelines of the Journal/Conference/Publisher.
- For thesis and other Institute documents, follow the guidelines:
  - If you are using AI tools to generate new content such as text, images, tables, code, etc. you must disclose their use (with details) either in the Methods or in the Acknowledgements section.
  - If entire sections are generated by AI tools, you must include details such as the prompt, version.
  - If you are using AI tools to edit and improve the quality of your existing text in much the same way you would use a typing assistant like Grammarly or a word processor, you need not disclose this use.

Private and Confidential information

- AI tools often use queries for their training, and hence may reveal their content in subsequent queries by other users.
- Hence, entering data into an AI tool should be regarded as equivalent to posting it in public (unless there are strong reasons to trust that a specific tool will not use your queries)
- Do not enter, contribute, or otherwise input sensitive, confidential, or restricted information into open AI tools.

Responsibility for Content

- AI tools are prone to errors and fabrications.
- If you use such systems in your work, as the author you are **wholly** responsible for the correctness of the content generated by the system.
- Another danger when using an AI system is that its output depends on its training data, which includes content that has been written by others who need to be credited.
- So, using the output could amount to plagiarism.
- Once more, as the author, you are **wholly** responsible for avoiding plagiarism.
- As AI systems cannot take responsibility, they **cannot** be authors.

Rules specific to programs, departments, courses

- Programs, departments, and courses may have specific rules for the use of AI tools.
- These must be followed in addition to the above rules.
- Those making rules should be aware that at present software detecting use of AI **does not work** and it has been observed that humans are also prone to errors and biases.

Guidelines and Best Practices for Teaching

Recommendations of the Committee

Positive Impacts on Teaching

- AI tools can provide a personalised learning experience to students. Students with diverse learning abilities and educational backgrounds could benefit. Students can explore topics, clarify doubts, and generate practice problems.
- Students could be challenged to solve more complex problems with these tools.
- AI tools can provide language assistance and help students communicate their ideas more clearly in course assignments.
- For teachers, AI tools can help with tasks, such as generating course plans and course policies, facilitating grading, and providing assistance to students outside class

Negative Impacts on Teaching

- Over-reliance on AI tools can reduce critical thinking, problem-solving skills, creativity, and a deep understanding of complex concepts.
- Responses from AI tools can be inaccurate and misleading.
- Output may be biased.
- There may be privacy, confidentiality, and plagiarism concerns.

Recommendations for Teaching (1 of 2)

- AI tools are no substitute for faculty and teaching assistants: For our courses to remain student-centric, instructors and TAs should continue to interact closely with students and devote personal attention to their progress.
- Course instructors may need to regularly update learning outcomes, pedagogy, and student assessments.
- Course assessments should adopt proactive and creative approaches, assuming that students will use AI tools.
- The institute could provide infrastructure and administrative support to help students and faculty keep up with changing technology and be informed on ethical and constructive use.

Recommendations for Teaching (2 of 2)

- We recommend transparent and simple course policies that should be communicated regularly.
  - Course policies related to the use of AI fall into one of four broad categories: prohibited, allowed with attribution, encouraged, or mandatory.
  - If a course does not have an explicit policy, we recommend **allowed with attribution** as the default Institution policy.
- The Student Handbook should include AI tools when describing plagiarism and correct attribution in the academic integrity section
- The roles and capabilities of AI tools are guaranteed to change rapidly. Therefore, it is essential that the policies be re-evaluated periodically while avoiding reactionary/impulsive stances.

Summary

- **AI** tools have come here to stay. They have positive effects as well as negative effects for education, learning, and research
- IISc faculty members recognize this and are in favour of responsible use of **AI** tools for maximizing the positives and minimizing the negatives.
- IISc students clearly want to use the **AI** tools and are looking forward to guidelines and training
- The committee has identified broad guidelines for creative, constructive, and responsible uses of **AI** tools for teaching and research
- Next step is to translate these guidelines into clear and transparent policies providing enough freedom for faculty members to fine-tune
- It is important to make suitable modifications to IISc documents, provide wide publicity, launch training modules to sensitize faculty as well as students, and to review the policies periodically, perhaps every year. Dept Chairs and faculty members will have a vital role to play.

IISc Faculty Survey: A Summary

IISc Faculty Survey (1 of 3)

- 97 responses
- Faculty were most familiar with AI tools for improving readability, summarising complex passages, and writing code
- Mixed responses on the impact of AI tools on teaching. Only a small fraction felt that it has only negative impacts on student learning
- Student use of AI tools could potentially affect course grades, as unproctored work contributes a large fraction to course grades

Clear Policies Needed (2 of 3)

- Vast majority felt a need for an up-to-date resource page for faculty
- Mixed responses on preferred policy for use of AI tools by students for course work
- Mixed responses on what constitutes appropriate use of AI tools by students
- Faculty vary in how they tackle the challenges posed by AI tools in teaching. Many indicated switching from unproctored, take-home assignments to proctored exams. Others have modified their assignments and tests so that AI tools are unlikely to affect assessment. Some see a positive use: more advanced assignments could be set now that students have access to AI tools

Faculty Survey: Summary (3 of 3)

The overall consensus for use in research and coursework:

Let us not completely restrict the use of these tools, as that would be counterproductive. Instead, let us hold training sessions for students on productive and ethical use versus inappropriate use of these tools.

IISc Student Survey: A Summary

IISc Student Survey (1 of 3)

- 525 responses
- The vast majority (53%) were familiar with ChatGPT and had already used it in coursework
- Students were most familiar with AI tools for improving readability, summarising research articles, and writing code. Relatively few used these tools for mathematical proofs or problem-solving.
- Most students identified coding and technical writing as key areas where AI tools could have a positive impact. Mathematics and problem solving were identified as the areas where AI tools were likely to be least helpful.
- Most students (over 60%) sought to use AI tools to improve the quality of their written submissions.

IISc Student Survey (2 of 3)

- The majority of students felt that the institute should provide clear guidelines.
- Over 95% felt that the institute should allow the use of AI tools in research and teaching.
- Students were divided in what constituted appropriate use for coursework: a large fraction (40%) sought to allow free usage but with attribution, while some felt it was appropriate for only specific tasks (e.g. writing, 26%).
- Student views on the use of AI tools in research was different from their views on use in coursework. Here, the majority opinion (39%) sought to restrict its usage for specific tasks (e.g. writing) whereas a smaller proportion (33%) sought usage with attribution. Only a minority (21%) sought to allow unrestricted usage.

IISc Student Survey (3 of 3)

- Overall, students consider that the benefits of AI tools outweigh their drawbacks, as long as the responsible use of AI tools is encouraged.
- Students felt that banning AI tools may place us at a disadvantage over other institutions that permit these tools. Some recommended that the University have a standing policy to purchase and make accessible the most recent AI tools.
- Students also advocated periodic reviews and revisions of usage policies.