LLMs in Education

Sowmya Vajjala
National Research Council, Canada
Applications of Language Models, Spring 2025, IIIT-H

AI and Education

- Learner support tools (e.g., personalized tutoring, feedback etc)

- Teacher support tools (e.g., for assessment, test creation etc)

- Learner Analytics (e.g., measuring learning outcomes)

- Digital Classrooms (e.g., interactive learning tools)

Not a new area of interest

AI in Education, Educational Data Mining, Learning @ Scale, Learning Analytics Conference, SIG-EDU @ ACL, Computer Assisted Language Learning etc are all some research communities that have been looking into how AI (+ technology in general) can be used in Educational setups - by/for teachers, students, and classrooms.

LLMs and Generative AI put the focus on use cases where AI was previously thought of as not good enough for the task, improving use cases where AI already made good progress, and enabling new use cases as well.

Today

- I will talk more from a language perspective i.e., LLM applications in Education involving language use somehow (e.g., reading, writing, speaking, listening)

- Most content is derived from a recent (Jan 2025) 3 hour tutorial we did on the same topic at COLING 2025 (https://coling2025-edu-llms.github.io/)

Language, Technology and Education

- Assistive tech (e.g., reading support, writing support, speaking support etc)

- Automated Assessment (e.g., content assessment, language assessment etc)

- Personalized Learning (e.g., feedback generation)

- Content Generation (e.g., generating lessons and lesson plans)

- Accessibility and Diversity (Supporting different groups of learners, different languages etc).

Today, we have both general purpose apps (e.g., DuoLingo, Khanmigo) as well as more specialized Intelligent Tutoring Systems (e.g., Mathia, Squirrel etc.).

LLMs and Reading/Listening

Potential Use cases:

- Help with word level stuff (hovering over words for meaning, pronunciation, etc.)
- Simplifying the language used in a text (word replacement, paraphrasing, etc.)
- Personalized text recommendations that suit the reading level and interests
- Generating content suitable for a reader's profile
- Generation of reading comprehension exercises to test understanding
- Assessing reading comprehension
- Summarizing longer texts
- Text to speech, read-aloud, etc. (accessibility)

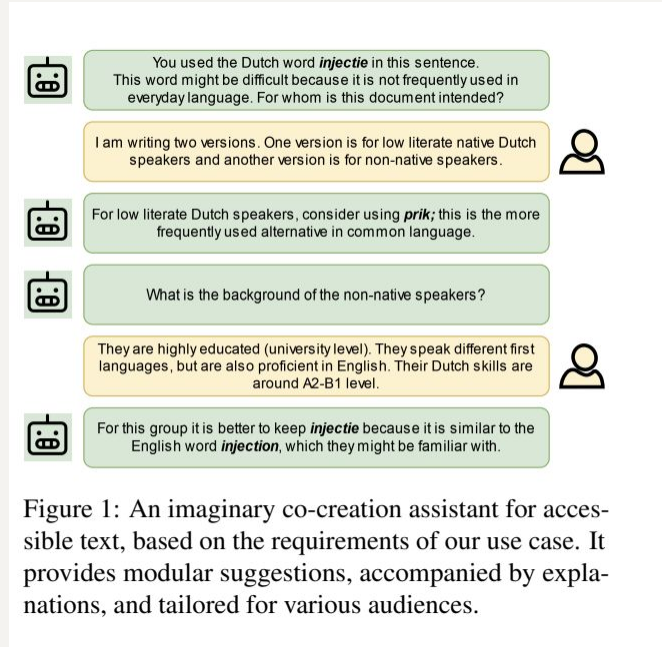Etc.

AI Assistants for Educational Content Creation

Figure 1: An imaginary co-creation assistant for accessible text, based on the requirements of our use case. It provides modular suggestions, accompanied by explanations, and tailored for various audiences.
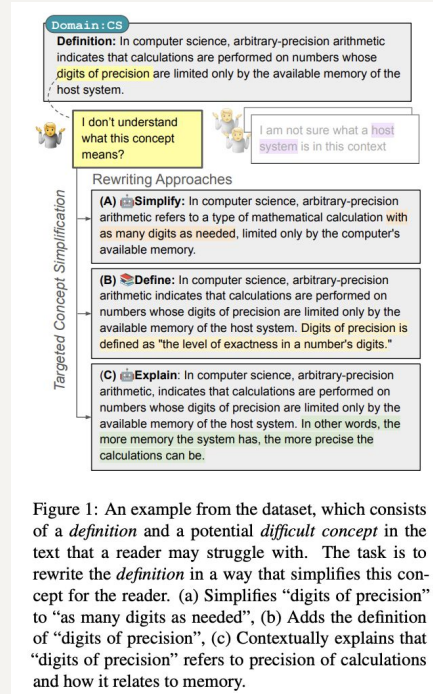
Kim et.al., 2024

Figure 1: An example from the dataset, which consists of a *definition* and a potential *difficult concept* in the text that a reader may struggle with. The task is to rewrite the *definition* in a way that simplifies this concept for the reader. (a) Simplifies "digits of precision" to "as many digits as needed", (b) Adds the definition of "digits of precision", (c) Contextually explains that "digits of precision" refers to precision of calculations and how it relates to memory.

Asthana et.al. 2024

Challenges and Opportunities

- We can quantify a "reading level" for a text, but can we point to specific parts of the text that are difficult to understand for a target reader?

- Can we personalize this kind of detection?

- How do we model the conceptual difficulty along with difficulty in form?

- What should we do about multimodal interactions while reading?

- What sort of datasets do we need? Can we build them easily, across languages?

LLMs and Writing

Prompting for:

- Writing assistance (paraphrasing, making things more formal or informal etc)

- Offering corrections (correcting spelling/grammar)

- Offering more explicit, personalized feedback on writing

- Assessment (grading of essays or short answers, for example)

Is fine-tuning necessary for such cases? Why? Why not?

Building End-to-End Writing Assistants
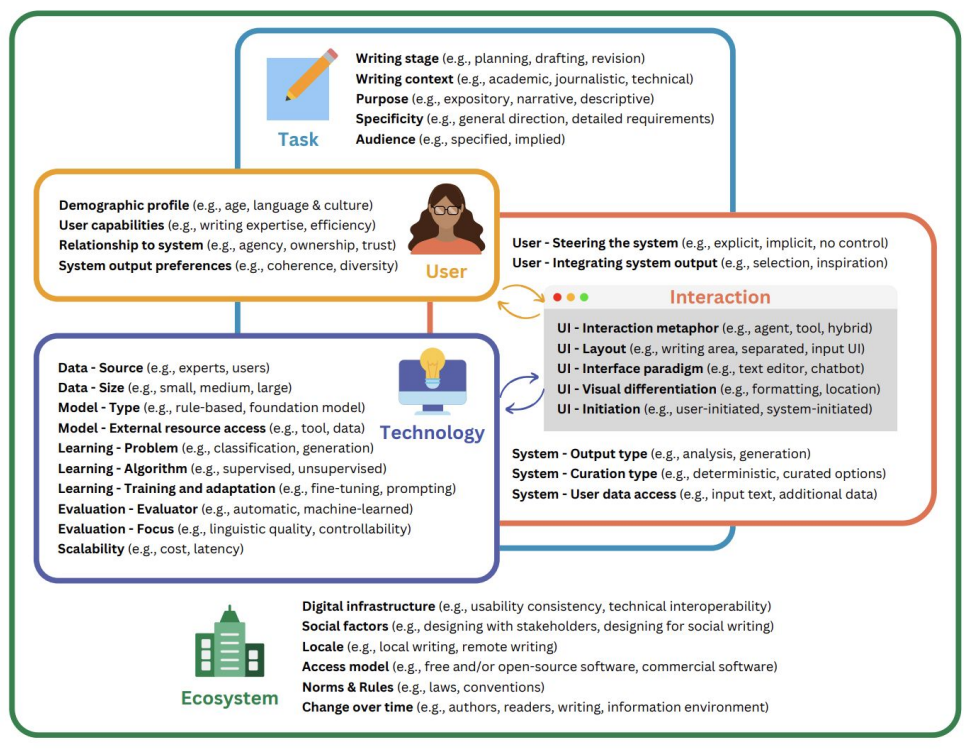
From Grammarly, CHI 2024

Raheja et.al. (2024)

Challenges and Opportunities

- Evaluation is challenging for some use cases such as correcting grammatical errors
- Offering detailed, personalized feedback is a new, but impactful and evolving area
- We *still* don't have much work on multilingual aspects of writing support/assessment, but LLMs may lower the entry barrier for new languages
- Automated Assessment with or without LLMs is in general controversial even in 2025.
- **One idea for dataset construction: Start with LLM generated output and then do manual labeling/corrections, instead of building datasets from scratch**?

LLMs and Speaking

Potential use cases:

- Spoken language proficiency assessment

- Feedback on the language form (grammar, vocabulary etc) based on speech

- Speaking practice (exercise generation + evaluation)

- Pronunciation training and practice

…

(b) Functional System (Web Application).

Kopparapu et.al., 2024
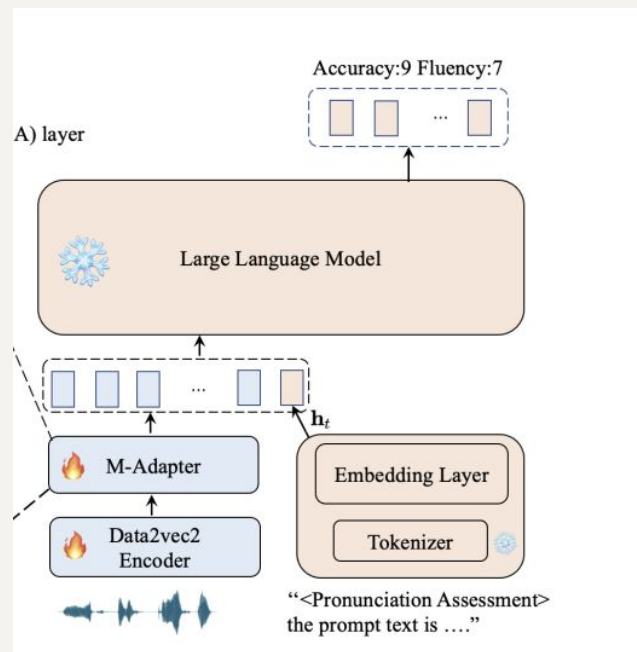
Fu et.al., 2024

Challenges and Opportunities

- Datasets that can be useful for these kind of use cases don't exist.

- Handling speech-text interaction, where relevant, is not easy

- Scaling across languages, accents etc is a big challenge.

- Feedback on form (i.e., language pronunciation, fluency etc) versus content (verifying the correctness of information in a spoken response) is unexplored.

- End-to-end speech interfaces (instead of intermediate text analysis stages) is another open possibility.

..

LLMs and Tutoring Systems

- Personalized (one-on-one) tutoring results in an improvement in learning gains with the effect size of **2 standard deviations** ([Bloom, 1984](#))

- One-on-one personalized tutoring is **not available** to everyone

- This is where technology (Intelligent Tutoring Systems, **ITS**) can help

- ITS that mimic personalized human tutoring in a computer-based environment are a **viable alternative** to human tutors (especially when in-person teaching is not possible).

Teaching LLMs Pedagogical Values

- LLMs, in their default form, *do not imitate* human tutoring behavior or exhibit pedagogy. What can we do?
- **Prompting:** adjust LLM behaviour for its response
- **Fine-tuning**
  - *Supervised fine-tuning:* train with pedagogically rich datasets
  - *RLHF:* optimize using human feedback

Problem Statement:
James writes a 3-page letter to 2 different friends twice a week. How many pages does he write a year?

Correct Answer:
He writes each friend 3*2=6 pages a week. So he writes 6*2=12 pages every week. That means he writes 12*52=624 pages a year.

$$3 * 2 * 2 * 52 = 624$$

Student Solution / Confusion:
Calculate the number of letters written in a week: 3 pages x 2 letters = 6 pages per week. Calculate the number of letters written in a year: 6 pages per week x 52 weeks = 312 pages per year.

$$3 * 2 * 52 = 312$$

Teacher: Hi Winnie, could you please walk me through your solution?
Student: Sure! I first calculated the number of letters written in a week, which is 3 pages x 2 letters = 6 pages per week. Then I multiplied that number by 52 weeks to get the number of pages written in a year, which is 312 pages per year.

Teacher: That sounds correct. Great work! **Factually incorrect** **Instruct-GPT**

Teacher: You calculated the number of pages written per week correctly. However, James writes two letters per week. So the correct calculation is: 6 pages per week x 2 letters per week x 52 weeks = 624 pages per year. Does that make sense? **Leaking solution** **ChatGPT**

Teacher: There is one important keyword there: twice. What does it refer to?

Source: Macina et al. (2023).

Challenges and Opportunities

- Can we build ITS that can handle **multiple subjects**?

- Can we extend them to **multiple languages**?

- Can they be easily adapted to **different learner levels**?

- Can **multiple modalities** be easily integrated?

- **Large-scale evaluation** with all stakeholders considered is still rare

Some General Challenges

- **Scalability**: Constrained by manual effort before, data availability today.

- **Personalization** can only be achieved with high-quality data
- **Cross-domain, cross-level**, cross-[...] applicability is still an issue

- **Technical challenges**: Factuality checks, Performance Issues etc.

- **Usability challenges**: Technology readiness, User's expertise.

- **Ethical challenges**: Transparency, Privacy, Equity, Cheating, Plagiarism etc.

Summary

- Vast space of ideas and possibilities with Generative AI

- There is a large existing body of work involving NLP, ML/DL etc, but LLMs open a whole range of new possibilities.

- But, there are also many challenges to deal with in terms of datasets, evaluation methodologies, addressing other domain specific challenges (e.g., ethics and integrity when using genAI in classrooms and for exams)

Many other topics I did not touch

- Building learner models - i.e., some kind of an evolving representation of what a learner has learnt, where they struggle, where they do well etc.

- Building other analytics e.g., summarizing student discussions/feedback, measuring engagement etc

- How student learning interacts with background knowledge, motivation etc.

… and so on.

Resources

- We compiled a set of NLP focused resources here: https://coling2025-edu-llms.github.io/

- [Generative AI and Applied Linguistics](#) - an introductory write-up I wrote for a Japanese journal (does not assume technical background, but assumes familiarity with language learning technologies)

- Visit the conference/journal pages of AIED, Educational Data Mining, Learning at Scale, Learning Analytics Conference etc for work outside NLP using LLMs

Questions?