Extended Abstract for the
Symposium on Integrating AI and Data Science into School Education Across Disciplines 2025

CONSTRUCTING & DECONSTRUCTING LARGE LANGUAGE MODELS IN HIGH SCHOOL CLASSROOMS

Karl-Emil Kjær Bilstrup
Aarhus University, Denmark, keb@cs.au.dk

Focus Topics: Tools, AI and Data Science Competencies

The sudden public availability and rapid adoption of large language models (LLMs), e.g., through OpenAI's services and their publicly available ChatGPT, will significantly impact future life and societies. LLM tools have rapidly been woven into the fabric of high school classrooms through commercial, publicly available tools that every student and teacher can access. This has created a push for teachers, school administrators, educational content developers, etc., to address how these technologies should play a role in the classroom, which has resulted in initiatives to explore the learning potentials of using these AI tools and the development of tools designed explicitly for Danish classrooms, such as SkoleGPT (school GPT). While this work, addressing how the teacher profession can handle these new technologies, is exciting and vital work, it misses crucial perspectives on how LLM technologies are also disrupting personal lives, most professions, and our societies. Studies find that 80% of the US workforce will be affected and that 15% of all work tasks can be completed significantly faster with LLM-powered technologies (Eloundou et al., 2023), and we are currently experiencing how LLM technologies are being implemented in various software products; from productivity tools in Microsoft Office packages to AI friends in SnapChat. I argue that students must be taught the skills and insights to navigate this changed world. Here, it is insufficient to have been taught by an AI tutor or know how to prompt a Chatbot to take the role of a historical character; a more fundamental understanding of the technologies and a more progressive attitude towards using them is required. This perspective builds on decades of efforts into teaching computational skills to support students in building, customizing, and taking advantage of computational technologies in their professional and personal lives (Iversen et al., 2018; Van Mechelen et al., 2022; Wing, 2006).

Currently, we see how commercial actors oversell LLMs as the idea of general intelligence (Siddarth et al., 2021) and how the public debate becomes centered around the prospects of such idea, instead of focusing on the implications of the current technology. There are already examples of how AI in education amplifies social biases and discriminates minority students (Selwyn, 2022). These issues will only be magnified with the rapid adoption of LLMs in classrooms. Instead, students should learn about LLMs as bounded mathematical systems that conduct algorithmic forecasting (Selwyn, 2022); how they are trained on vast amounts of data from the web that mainly expresses values expressed through English language (Brown et al., 2020); how climate impact of LLMs has become a serious issue given the enormous amount of energy that is required to train the computational models (Hershcovich et al., 2022; Selwyn, 2022); and how they are fine-tuned through manual human labor (D. Wang et al., 2022).

The recent success of LLMs is among other innovative steps based on aligning the models through human feedback processes where humans manually annotate data and provide feedback on LLMs' outputs (Y. Wang et al., 2023). The human-annotated data are used to align foundational LLMs (models trained on data scraped from the internet, books, etc.) to human preferences and to train them in solving specific tasks in a process called fine-tuning. Experts can conduct the data annotation if the goal is to fine-tune a model to solve specific tasks in a research field or a company context. But for more general tasks, such as acting as a helpful chatbot that does not hallucinate or make discriminating utterances, low-wage labor in the global south is primarily used (The Washington Post, 2023; D. Wang et al., 2022). This process poses issues with transparency and how the plurality of local human cultures and values are taken into account (Johnson et al., 2022). Studies have shown how LLMs demonstrate social biases, favor more progressive political views than the general public (Chang et al., 2024), and mainly align with dominant US public opinions (Johnson et al., 2022). This causes urgent social and democratic issues when recent studies demonstrate how LLMs impact how we communicate and form social relationships through text (Hohenstein et al., 2023) and how users of LMMs unknowingly adopt

In S. Podworny & S. Schönbrodt (Eds.), *Towards Fostering AI and Data Science Literacy in Schools Across Disciplines. Proceedings of the 1st Symposium on Integrating AI and Data Science into School Education Across Disciplines (AIDEA1 2025)*, Salzburg, Austria.

opinions embedded in LLMs – not just in the text they produce but also when asked about their opinion afterwards (Jakesch et al., 2023).

In this project, we will engage Math and Informatic teachers and students in hands-on activities of fine-tuning LLMs to students' preferences and solving high school subject-specific tasks. To address challenges about the scale of datasets, computing, and hosting required to conduct such activities, this project takes a crowdsourcing approach where teachers and students contribute to shared data sets among multiple classrooms, which we will use to fine-tune LLMs that we host to make them available to the same teachers and students. Students will be engaged in activities such as using subject knowledge to create exemplary solutions through generating or annotating text, measuring agreement among annotators, formal evaluation of fine-tuned models, and using the fine-tuned models in analytical activities.

References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are fewshot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), Advances in neural information processing systems (pp. 1877–1901, Vol. 33). Curran Associates, Inc. https://proceedings.-neurips.cc/paperfiles/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64aPaper.pdf

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models ACM Trans. Intell. Syst. Technol. https://doi.org/10.1145/3641289

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. https://arxiv.org/abs/2303.10130

Hershcovich, D., Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2022). Towards climate awareness in nlp research. https://arxiv.org/abs/2205.05071

Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. Scientific Reports, 13 (1), 5487.

Iversen, O. S., Smith, R. C., & Dindler, C. (2018). From computational thinking to computational empowerment: A 21st century pd agenda. https://doi.org/10.1145/3210586.3210592

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., & Naaman, M. (2023). Co-writing with opinionated language models affects users' views. https://doi.org/10.1145/3544548.3581196

Johnson, R. L., Pistilli, G., Men´edez-Gonz´alez, N., Duran, L. D. D., Panai, E., Kalpokiene, J., & Bertulfo, D. J. (2022). The ghost in the machine has an american accent: Value conflict in gpt-3.

Selwyn, N. (2022). The future of ai and education: Some cautionary notes. European Journal of Education, 57 (4), 620–631. https://doi.org/https://doi.org/10.1111/ejed.12532

Siddarth, D., Acemoglu, D., Allen, D., Crawford, K., Evans, J., Jordan, M., & Weyl, E. (2021). How ai fails us. arXiv preprint arXiv:2201.04200.

The Washington Post. (2023). Behind the ai boom, an army of overseas workers in 'digital sweatshops'. Retrieved January 16, 2024, from https://www.washingtonpost.com/world/2023/08/28/scaleai-remotasks-philippines-artificial-intelligence/

Van Mechelen, M., Smith, R. C., Schaper, M.-M., Tamashiro, M. A., Bilstrup, K.-E. K., Lunding, M. S., Petersen, M. G., & Iversen, O. S. (2022). Emerging technologies in k–12 education: A future hci research agenda. ACM Trans. Comput.-Hum. Interact. https://doi.org/10.1145/3569897

Wang, D., Prabhat, S., & Sambasivan, N. (2022). Whose ai dream? in search of the aspiration in data annotation. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3491102.3502121

Wang, Y., Zhong, W., Li, L., Mi, F., Zeng, X., Huang, W., Shang, L., Jiang, X., & Liu, Q. (2023). Aligning large language models with human: A survey.

Wing, J. M. (2006). Computational thinking. Commun. ACM, 49 (3), 33–35. https://doi.org/10.1145/1118178.1118215