The Unseen A+ Student: Navigating the Impact of Large Language Models in the Classroom

Matyáš Boháček [1]

Abstract

The recent boom of so-called generative artificial intelligence (AI) applications, namely large language models such as ChatGPT, took the public discourse by storm, disrupting many fields and industries. Education, being one of them, was left with little to no information about the impacts of such an unprecedented technology, let alone holistic methodologies on how to reflect and adopt it. Therefore, many educators rush to adopt policies banning the technology but find themselves incapable of reliably detecting the cheaters. All in all, the distribution established chaos in the education community. To address this problem, we propose a methodology assessing relevant information for educational institutions that can guide their decision-making around generative AI practices and policies. We demonstrate this methodology in a Czech high school, conduct a student assessment study, and make the dataset available to the public.

1. Introduction

When OpenAI introduced ChatGPT[1] in November of 2022, millions of people worldwide could suddenly utilize the power of large language models (LLMs) in an intuitive, chat-like user interface. Its popularity skyrocketed, and we saw enthusiasts — experts and laypeople alike — hunt for optimal prompts, create various automation pipelines, and share their discoveries online.

It is not hard to see why so many people have fallen for it (Haque et al., 2022): type in a message (or a command), just as you would to a real human assistant, and you will not be disappointed with the result. ChatGPT can write an email, summarize text, prepare notes, brainstorm ideas–but most importantly, save time (Nuzula & Amri, 2023; Joublin et al., 2023; Li et al., 2023). On the backend, everything is orchestrated by a neural network, a form of a machine learning (ML) model trained on large-scale data from the internet (Lu et al., 2023). Nonetheless, recent discourse includes it under the shortcut umbrella term of artificial intelligence (AI).

Hand in hand with the hype and excitement came worries about how such a powerful technology could be misused, prominently in education. OpenAI benchmarked the off-the-shelf ChatGPT with GPT-4 on numerous academic exams and found that it performs well above average human students in many subjects (OpenAI, 2023). In SAT, the standardized test for American college applications, the model achieved the 93rd and 89th percentile on Evidence-Based Reading & Writing, and Math parts, respectively. In both the Advanced Placement (AP) Art History and Biology Exams, it got 5, the highest score.

Educational institutions recently began to respond and introduce their policies on the use of this technology. While some educators and organizations pioneer frameworks to include AI in the classroom and plan to experiment with different approaches in the upcoming months (Wood & Kelly, 2023), many have strictly prohibited it, including College Board (Board), which runs SAT and AP exams. Many high schools and universities soon followed (Cassidy, 2023; Yang, 2023; Jimenez, 2023). Jointly, they implemented detectors of AI-generated texts, which should, similarly to plagiarism detectors, spot the cheaters (Lonas, 2023; Hsu, 2023). However, unlike plain plagiarism, proving that students used an AI model to generate their text is significantly more complex and prone to false positive findings (Sadasivan et al., 2023).

Amidst this rapid development and change in school policies, many questions remain unsolved. Therefore, we interviewed 4 high school teachers in Prague, Czech Republic, to understand what information they would need to adapt their teaching to unprecedented access to AI-driven text

This is an adapted paper, first presented at AIED-W 2023 *Workshop on Empowering Education with LLMs*, intended for presentation at ICML-W 2023 *Challenges of Deploying Generative AI*. [1]Gymnasium of Johannes Kepler, Parléřova 2/118, Prague, 169 00, Czech Republic. Correspondence to: Matyáš Boháček <matyas.bohacek@matsworld.io>.

Workshop on Challenges in Deployable Generative AI at International Conference on Machine Learning (ICML), Honolulu, Hawaii, USA. 2023. Copyright 2023 by the author(s).

[1]https://openai.com/blog/chatgpt

generation.

As some of them pointed out, OpenAI's report (OpenAI, 2023) includes mostly one kind of school assessment (exams) in one language (English). Above all, they need to understand how well the system performs in Czech to take any action. If its qualities are applicable to Czech school assignments, they want to see which kinds of tasks it can solve and how it compares to human students. Lastly, they need to know if such AI-generated responses are identifiable by humans or an automated system.

We developed a holistic study for education environments to answer these questions and ran it at the school. While tailored to the school's context and needs, we believe similar efforts are transferable to other regions, languages, and types of schools.

2. Related Work

Recently, the literature has begun exploring the implications of widely accessible AI tools for education. One of their fundamental premises is that they will enable personalized and interactive learning, with tailored instructions and more continuous evaluation (Baidoo-Anu & Ansah, 2023). Moreover, they are expected to accelerate students' research and writing process, allowing for more analytical and collaborative activities (Adiguzel et al., 2023). Some studies also focus on how AI and LLMs could benefit specific subjects, most prominently medicine (Sallam, 2023).

On the other hand, many recent works outline the potential dangers AI and LLMs pose for education. Megahed et al. (2023) show that ChatGPT struggles with nuanced tasks, such as explaining less widely known terms or creating factual content from scratch, and thus may be untrustworthy when teaching new content. Rahman & Watanobe (2023) describe specific misuses (e.g., cheating on online exams or generating essay assignments) and hypothesize that over-reliance on AI could eventually diminish critical thinking skills.

Many recent works studied whether humans can distinguish LLM-generated and human-produced texts (Dugan et al., 2023; Clark et al., 2021). The results suggest that — in most contexts — human judgment is no better than guessing on this task. However, the identification accuracy slightly improves with training on which patterns of LLM-generated text to observe.

With poor human accuracy, different automatic approaches to distinguish AI- and human-produced text have been introduced (Jawahar et al., 2020; Ippolito et al., 2019; Gehrmann et al., 2019; Crothers et al., 2022). Nevertheless, their precision varies significantly given the context and usually requires the knowledge of the LLM architecture used for the generation in the first place, limiting their practical use. Additional limitations — including the bias of these systems against non-native English writers — have been identified (Liang et al., 2023).

As for employing AI detectors in educational contexts, some opinion pieces have suggested that their reliability may be problematic depending on the context (Alimardani & A. Jane, 2023); nonetheless, to the best of our knowledge, there are no systematic analyses of this phenomenon to date.

3. Dataset

To answer the teachers' questions, we collected a dataset of coursework from the school. All of the assignments were completed in years 2019-2023. With many different kinds of written assignments, we divided the dataset into 2 primary parts and 5 latter sub-splits. For every ChatGPT generation (initially performed atop GPT 4.0 backbone), we generate GPT 3.5 and 3.5 Legacy backbone-driven counterparts, resulting in 3 variants of the synthesized text. We include a complete set of our prompts in Appendix B.

3.1. Long-form Theses

We first assemble 20 final high school theses: 10 for the subject of 'Czech Language and Literature' and 10 for 'Humanities'. Each work was written in Czech, consists of some 30 to 60 pages, and follows the general guidelines of formal academic writing. On top of these, we create 2 sub-splits, each holding an equal ratio of data from both subjects.

Sub-split A: holds abstract and keyword pairs for 10 theses. We generated the 3 synthetic alternative abstracts and keywords by including the introduction and conclusion of the respective work in the prompt.

Sub-split B: holds two subsequent paragraphs of text, with 3 synthetic alternatives that replace the second paragraph.

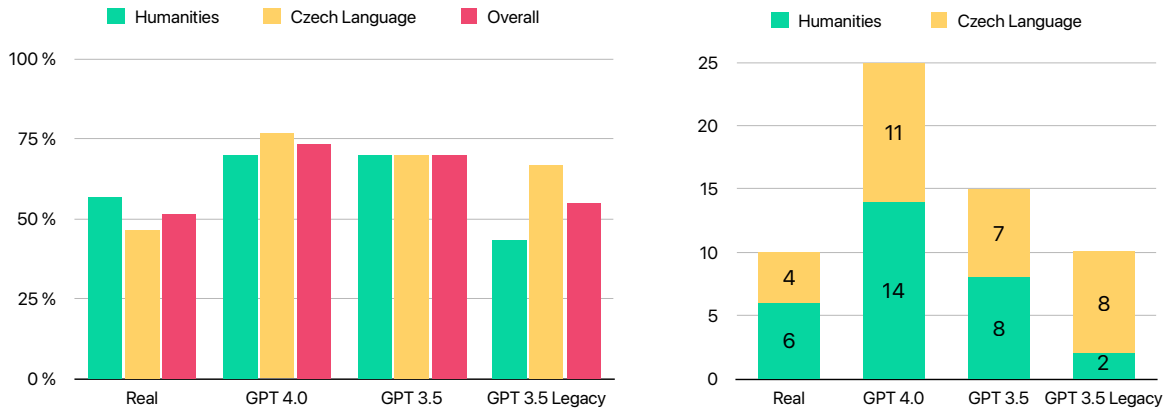
3.2. Short Assignments

Next, we assemble various assignments from different subjects. For each assignment, we include 10 human-written responses and generate 3 alternatives using ChatGPT, only given the instructions (i.e., we did not present the system with students' work).

Sub-split C: holds the instructions and responses of an essay assignment in a 'English as the Second Language' course.

Sub-split D: holds the instructions and responses of an essay assignment in a 'German as the Third Language' course.

Sub-split E: holds the instructions and responses of a quiz assignment in a 'Math' class.



(a) Proportion of instances in which the abstract and keyword pairs were selected as relevant (given the scope of the work).

(b) Absolute instances in which the abstract and keyword pairs were selected as the best, given the selection.

Figure 1. Results of the peer quality assessment conducted on the sub-split A. The results are reported for authentic texts and three versions of ChatGPT (4.0, 3.5, and 3.5 Legacy). For each source, the scores are divided into subjects (and averaged under 'Overall', if relevant).

4. Human Assessment

We recruited 6 student peers from the school, ages 18-20. Each participant was instructed on the task and later presented with the same data (i.e., the set of questions and reference texts was identical for each participant). We present the set of instructions and questions in Appendix C. Given average reading speeds, we designed the overall annotation task to take 75 minutes.

4.1. Quality Assessment

First, we assessed how the generated and authentic abstracts compare in terms of relevance (by peer student measures). For all 10 theses in sub-split A, the participants were presented with 4 alternative abstracts and keywords (1 authentic, 3 generated). We did not disclose which one is authentic and which is generated. The participants then had to select all options they deemed relevant (i.e., meeting the formal criteria and corresponding to the topic) and then select the single best one.

Shown in Figure 1a are the proportions of abstracts selected as relevant, grouped by model version and subject (the 'Overall' bar averages the subject-specific scores). Shown in Figure 1b are the absolute instances selected as the single best variants in the given selection, grouped by model version and subject.

We found that, on average, participants ranked abstracts generated by ChatGPT 3.5 Legacy similarly to the authentic ones, with around 50% of instances deemed relevant. Abstracts generated with ChatGPT 4.0 and 3.5 were perceived noticeably better: nearly 75% of their instances were deemed relevant.

As for the best option selection task, texts from ChatGPT 4.0 dominated, with a total of 25 of its instances selected as the best option. GPT 3.5 texts ranked second with 15 instances; authentic and GPT 3.5 Legacy texts share the last rank with 10 instances. Overall, there seems to be little to no statistically significant difference between the observed subjects.

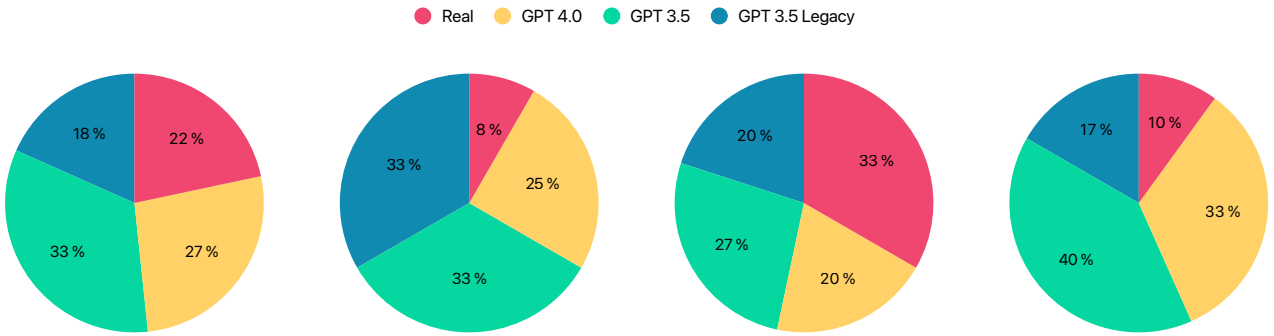
4.2. AI Text Identification

Next, we assessed whether participants could identify the authentic continuation of texts from sub-split B. Given 4 options, they were tasked to select the 1 authentic text among 3 generated ones. In general, humans without prior briefing on how to spot AI text are not able to do so (Dugan et al., 2023; Clark et al., 2021); we were interested in whether this translates to the educational paradigm.

Shown in Figure 2a is the overall distribution of texts identified as authentic, grouped by the origin (e.g., authentic or model type). Authentic texts were selected as such only 22% of the time, which suggests that the participants are more likely to identify generated texts as authentic.

Most continuations in sub-split B (8 of the 10) were just a paragraph long. We wondered if an extended generation range would affect the participants' judgment and created 2 special cases, where the continuation spans 3 paragraphs. Figure 2b captures the ranking distribution for this subcase. Interestingly, pro-longed authentic texts were even less likely to be deemed authentic compared to their prolonged counterparts.

Figures 2c and 2d divide the analysis by subjects: 'Humanities' and 'Czech Language & Literature', respectively. In



(a) Overall (b) Long Continuations (c) Humanities subject (d) Czech Language & Lit.

Figure 2. Results of the AI text identification assessment on the sub-split B. The results are reported for authentic texts and three versions of ChatGPT (4.0, 3.5, and 3.5 Legacy). Apart from overall statistics (a), results on a subset of longer continuations are presented (b), as well as subject-specific results for 'Humanities' and 'Czech Language and Literature' (c, d).

'Humanities', participants tend to select the authentic texts correctly more than the remaining classes but latter subject suffers from a dominance of the AI-generated texts.

5. Automatic Assessment

Lastly, we tested the following publicly available services, promising to identify texts generated using ChatGPT:

- **Content at Scale: AI Content Detector**[2], yielding a likelihood of the text being written by human;
- **GPTZero**[3], classifying human-written, mixed, and AI-written texts;
- **OpenAI's AI Text Classifier**[4], classifying very unlikely, unlikely, unclear, possibly, or likely AI-generated texts;
- **Writer: AI Content Detector**[5], yielding a likelihood of the text being written by human;
- **ZeroGPT**[6], yielding a likelihood of the text being written by AI.

Even though most of these services provide a nuanced assessment, we converted them to a binary classification for the purposes of our study. We do not report conventional metrics that would indicate the performance of individual tools, as they all completely failed our test. When evaluated on sub-set A, OpenAI's AI Text Classifier predicted that all the items are AI-generated, while the rest of the services classified all the items as human-produced. This means that, if used in practice, all students who wrote the material in our dataset – regardless of whether they used AI or not – would be classified as cheaters or rule-abiding students, depending on the service. This shows that current services cannot detect AI content in Czech, at least in the educational domain.

6. Conclusion

To summarize, we identified critical information for educators to adapt their teaching to publicly available generative AI systems. We outlined a methodology to quickly gather it, tailored to their regional and language context. To further demonstrate the methodology in action, we collected a dataset of authentic high school coursework, including both long-form theses and short assignments, from a public high school in the Czech Republic. We generated their AI alternatives and text continuations using ChatGPT with 4.0, 3.5, and 3.5 Legacy backbones. We make the data publicly available for open-domain research and analyses at https://www.matyasbohacek.com/topics/ai-education/.

Through a study involving student peers, we found that ChatGPT can quickly produce high-school-level coursework that peers consider better than human-written text, even in a low-resourced language like Czech. Moreover, we show that the AI text detectors, which are slowly rolling out to campuses and educational centers worldwide, fail to identify these texts in Czech.

Our methodology may serve as a reference for educational institutions that strive to understand the impact of AI tools in their specific context and take appropriate action. Moreover, it could also be adopted by AI service providers, which would release their findings at launch, preventing chaos and misconceptions and accelerating the response of the

[2] https://contentatscale.ai/ai-content-detector/
[3] https://gptzero.me/
[4] https://platform.openai.com/ai-text-classifier
[5] https://writer.com/ai-content-detector/
[6] https://www.zerogpt.com/

educational institutions.

Acknowledgements

We would hereby like to thank Dr. Činátlová for her valuable insight and initiative when communicating with teachers and students at the subject high school, as well as all her many thought-provoking comments. Additionally, we would like to thank Progresus TOGETHER foundation for their generous sponsorship of this research and mobility-associated costs.

References

Adiguzel, T., Kaya, M. H., and Cansu, F. K. Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, 2023.

Alimardani, A. and A. Jane, E. We pitted ChatGPT against tools for detecting ai-written text, and the results are troubling, Apr 2023. URL https://theconversation.com/we-pitted-chatgpt-against-tools-for-detecting-ai-written-text-and-the-results-are-troubling-199774.

Baidoo-Anu, D. and Ansah, L. O. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *SSRN Electronic Journal*, 2023.

Board, A. C. C. 2022-23 guidance for artificial intelligence tools and other services. URL https://apcentral.collegeboard.org/exam-administration-ordering-scores/administering-exams/preparing-for-exam-day/exam-security/artificial-intelligence-tools.

Cassidy, C. Australian universities split on using new tool to detect AI plagiarism, Apr 2023. URL https://www.theguardian.com/australia-news/2023/apr/16/australian-universities-split-on-using-new-tool-to-detect-ai-plagiarism.

Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., and Smith, N. A. All that's 'human' is not gold: Evaluating human evaluation of generated text. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 7282–7296, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.565. URL https://aclanthology.org/2021.acl-long.565.

Crothers, E., Japkowicz, N., and Viktor, H. L. Machine generated text: A comprehensive survey of threat models and detection methods. *ArXiv*, abs/2210.07321, 2022.

Dugan, L., Ippolito, D., Kirubarajan, A., Shi, S., and Callison-Burch, C. Real or fake text?: Investigating human ability to detect boundaries between human-written and machine-generated text. In *The 37th AAAI Conference on Artificial Intelligence*, 2023.

Gehrmann, S., Strobelt, H., and Rush, A. M. GLTR: Statistical detection and visualization of generated text. In *Annual Meeting of the Association for Computational Linguistics*, 2019.

Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., and Ahmad, H. "i think this is the most disruptive technology": Exploring sentiments of chatgpt early adopters using twitter data. *ArXiv*, abs/2212.05856, 2022.

Hsu, J. Plagiarism tool gets a ChatGPT detector – some schools don't want it, Apr 2023. URL https://www.newscientist.com/article/2367322-plagiarism-tool-gets-a-chatgpt-detector-some-schools-dont-want-it/.

Ippolito, D., Duckworth, D., Callison-Burch, C., and Eck, D. Automatic detection of generated text is easiest when humans are fooled. In *Annual Meeting of the Association for Computational Linguistics*, 2019.

Jawahar, G., Abdul-Mageed, M., and Lakshmanan, L. V. S. Automatic detection of machine generated text: A critical survey. In *International Conference on Computational Linguistics*, 2020.

Jimenez, K. "this shouldn't be a surprise" the education community shares mixed reactions to ChatGPT, Feb 2023. URL https://eu.usatoday.com/story/news/education/2023/01/30/chatgpt-going-banned-teachers-sound-alarm-new-ai-tech/11069593002/.

Joublin, F., Ceravola, A., Deigmoeller, J., Gienger, M., Franzius, M., and Eggert, J. A glimpse in chatgpt capabilities and its impact for ai research. *ArXiv*, abs/2305.06087, 2023.

Li, B., Fang, G., Yang, Y., Wang, Q., Ye, W., Zhao, W., and Zhang, S. Evaluating chatgpt's information extraction capabilities: An assessment of performance, explainability, calibration, and faithfulness. *arXiv preprint arXiv:2304.11633*, 2023.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., and Zou, J. Y. Gpt detectors are biased against non-native english writers. *ArXiv*, abs/2304.02819, 2023.

Lonas, L. Plagiarism finder Turnitin adds AI detection amid popularity of ChatGPT, Apr 2023. URL https://thehill.com/policy/technology/3928562-plagiarism-finder-turnitin-adds-ai-detection-amid-popularity-of-chatgpt/.

Lu, Q., Zhu, L., Xu, X., Xing, Z., and Whittle, J. Towards responsible ai in the era of chatgpt: A reference architecture for designing foundation model-based ai systems. *ArXiv*, abs/2304.11090, 2023.

Megahed, F. M., Chen, Y.-J., Ferris, J. A., Knoth, S., and Jones-Farmer, L. A. How generative AI models such as ChatGPT can be (mis)used in SPC practice, education, and research? an exploratory study. *ArXiv*, abs/2302.10916, 2023.

Nuzula, I. F. and Amri, M. M. Will chatgpt bring a new paradigm to hr world? a critical opinion article. *Journal of Management Studies and Development*, 2023.

OpenAI. GPT-4 technical report. *ArXiv*, abs/2303.08774, 2023.

Rahman, M. M. and Watanobe, Y. ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 2023.

Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., and Feizi, S. Can AI-generated text be reliably detected? *ArXiv*, abs/2303.11156, 2023.

Sallam, M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 11, 2023.

Wood, P. and Kelly, M. L. "everybody is cheating": Why this teacher has adopted an open ChatGPT policy, Jan 2023. URL https://www.npr.org/2023/01/26/1151499213/chatgpt-ai-education-cheating-classroom-wharton-school.

Yang, M. New York City schools ban AI chatbot that writes essays and answers prompts, Jan 2023. URL https://www.theguardian.com/us-news/2023/jan/06/new-york-city-schools-ban-ai-chatbot-chatgpt.

A. Dataset Samples

	ORIGINAL (CZECH)	TRANSLATED (ENGLISH)
	Název: Feminizace migrace **Předmět:** Humanitní studia	**Title:** Feminization of migration **Subject:** Humanities
ORIGINAL (HUMAN)	**Abstrakt:** Práce se zaměřuje na ženskou migraci a její specifika. V práci je popsáno, kterým okolnostem ženy při migraci čelí a je snaha upozornit na mýty a stereotypy, které kolem migrujících žen panují. **Klíčová slova:** migrace, ženská migrace, migrace v ČR, teorie push-pull, informativnost v migraci, care-drain, integrace migrantů, překvalifikovanost migrantů	**Abstracts:** This thesis focuses on female migration and its specifics. The thesis describes the circumstances that women face during migration and tries to highlight the myths and stereotypes that exist around women migrants. **Keywords:** Migration, female migration, migration in Czechia, push-pull theory, informativeness in migration, care-drain, integration of migrants, overqualification
AI-GENERATED (ChatGPT)	**Abstrakt:** Tato maturitní práce se zabývá feminizací migrace jako spojením dvou sociálně zranitelných skupin, žen a migrantů. Práce popisuje intenzitu feminizace migrace, zdrojové faktory, které ji podporují a konkrétní příklady feminizace migrace v České republice. **Klíčová slova:** feminizace migrace, ženská migrace, Česká republika, teoretické popisy, praktické fakty.	**Abstract:** This thesis explores the feminization of migration as the coming together of two socially vulnerable groups, women and migrants. The thesis describes the intensity of feminization of migration, the resource factors that support it and specific examples of feminization of migration in Czechia. **Keywords:** Feminization of migration, female migration, Czech Republic, theoretical descriptions, practical facts.

Figure A.1. Representative example of an item from the dataset's subsplit A, along with its metadata. The original human-written text is shown above one of its three alternatives generated using ChatGPT.

B. Prompts

Original (Czech)	Translated (English)
Pokračuj v psaní práce. Abstrakt: `[abstract]` Klíčová slova: `[keywords]` Úvod:	Resume writing of this thesis. Abstract: `[abstract]` Keywords: `[keywords]` Introduct:
Toto je odborná práce na téma "`[topic]`". Pokračuj v psaní textu: " `[portion of the text]`	This is a thesis concerning the topic of "`[topic]`". Resume writing of this thesis: `[portion of the text]`
Toto je úvod maturitní práce: "`[introduction]`" Toto je závěr maturitní práce: "`[conclusion]`" Napiš abstrakt ve stejném stylu:	This is the introduction of a high school leaving thesis: "`[introduction]`" This is the conclusion of a high school leaving thesis: "`[conclusion]`" Write an abstract in the same style:
Toto je úvod maturitní práce: "`[introduction]`" Toto je závěr maturitní práce: "`[conclusion]`" Napiš krátkou anotaci a klíčová slova:	This is the introduction of a high school leaving thesis: "`[introduction]`" This is the conclusion of a high school leaving thesis: "`[conclusion]`" Write a short annotation and keywords:
Toto je zadání úkolu do předmětu `[subject]` na střední škole: "`[instructions]`". Vypracuj úkol:	This is an assignment in `[subject]` class at a high school: "`[instructions]`". Complete the assignment:

Table B.1. Prompts used for generating the alternatives or continuations of coursework within our dataset.

C. Survey Instructions and Questions

Original (Czech)	**Translated (English)**
Pomocí tohoto dotazníku analyzujeme, zda jsou generativní AI modely schopné odpovídat na různé typy úkolů a zda jsou tyto texty rozpoznatelné od těch skutečných, lidsky napsaných.	With this questionnaire, we seek to analyze whether generative AI models are able to complete different kinds of coursework and whether these texts are recognizable from real, human-written ones.
Níže uvidíte několik verzí abstraktu ke stejné maturitní práci z humanitních studií nebo českého jazyka. U každé práce zodpovězte následující otázky: 1. Které z navrhovaných možností **fungují jako adekvátní abstrakt** (tzn. nastiňují předmět a cíl práce, krátce shrnují obsah, a hlavně navnazují čtenáře*řku k tomu, aby si celou práci přečetl*la)? — *můžete zvolit libovolný počet odpovědí (tzn. klidně všechny nebo žádnou)* 2. Která z navrhovaných možností je, podle Vás, **pro svůj účel nejvhodnější**? — *volte právě jednu možnost*	Below, you will be presented with different alternatives for an abstract to accompany graduation theses (from Humanities or Czech language subjects). For each thesis, answer the following questions: 1. Which suggested options **work as an adequate abstract** (i.e., outline the topic and aims of the work, briefly summarize its contents, and—perhaps most importantly—grasp the reader)? — *you may select any number of options (i.e., including all and none)* 2. Which of the proposed options do you think is the **most suitable for its purpose**? — *you must select only one option*
Které z navrhovaných možností fungují jako adekvátní abstrakt?	Which suggested options work as an adequate abstract?
Která z navrhovaných možností je, podle Vás, pro svůj účel nejvhodnější?	Which of the proposed options do you think is the most suitable for its purpose?
Níže uvidíte několik krátkých úryvků z maturitních prací z humanitních studií nebo českého jazyka. U každého se nachází 4 alternativní pokračování – 1 skutečné (původní), 3 vygenerována pomocí GPT-4. Vyberte vždy tu variantu, u níž si myslíte, že pochází z původní, člověkem psané práce.	Below, you will be presented with short excerpts from graduation theses (from Humanities or Czech language subjects). For each, there are 4 alternative continuations - 1 real (original) and 3 generated by GPT-4. For eacg thesis, select the variant you think comes from the original, human-written work.
Která z navrhovaných možností, podle Vás, pochází z původní, člověkem psané práce?	Which of the proposed options do you think comes from the original, human-written work?

Table C.2. Instructions and questions in the digital questionnaire participants completed after an in-person briefing.