Will AI Help or Hurt Learning? *

Catalina Franco[†] Natalie Irmert[‡] Siri Isaksson[§]

September 24, 2025

Please do not cite or circulate.

Abstract

AI is transforming how students learn. This development raises pressing concerns about whether it promotes equitable access to education—or widens existing learning gaps. Who will benefit from AI in their studies, and who may be left behind? We examine this question through a pre-registered lab experiment (N = 572) with two stages: a study phase, where participants learn a novel topic, followed by an exam on the same topic. Participants were randomly assigned to one of three conditions that determined what tools they could use while studying: (1) Control (browsing only), (2) AI-assisted (access to a generative AI), or (3) AI-guided (AI access with additional guidance). During the study phase, students could allocate their time between reading materials, solving practice questions, and using the assigned tools. All participants then completed an exam without AI access, allowing us to causally estimate the effects of AI on learning. First, we find no overall effect of AI access on learning outcomes. However, this average masks significant heterogeneity: high-GPA women clearly benefit from AI, while men show mixed responses. Second, we find that students with AI access attempt fewer practice questions, meaning that students with access to AI engage less in active problem solving. Finally, we analyze participants' prompt data to explore how students interact with AI and how this affects learning. Together, the results show that AI reshapes not only who benefits from learning, but how learning itself takes place.

*We are grateful for funding from the Jan Wallanders och Tom Hedelius stiftelse samt Tore Browaldhs stiftelse (grant P24-0189). We thank Iver Finne, and the seminar audiences at the 2025 ASSA Meeting in San Fransisco, IFN Stockholm, Umea University, NTNU Trondheim, 2025 ESA-Asia, ISER Osaka Workshop, 2025 AFSEE Meeting, 2025 SEET Meeting in Dijon, 2025 CEN in Aarhus, the 2025 Stockholm Uppsala Education Economics Workshop, and the 2025 CESifo / ifo Junior Workshop on the Economics of Education.

[†]Center for Applied Research (SNF) at NHH – Norwegian School of Economics.

[‡]Lund University

[§]Burgundy School of Business.

1 Introduction

The rapid advancement of generative AI is set to transform several spheres of society. Among these, education stands out, as AI tools are poised to drastically change how students acquire knowledge. Through AI, a student today can access PhD-level writing and analysis with a simple prompt, rather than reaching this level of understanding through their own effort over time. This raises a fundamental question about the technology's impact on human capital accumulation: *does AI access help or hurt learning?*

The impact of AI on student learning remains unclear and likely depends on *how* students use these tools. AI may enhance learning by complementing student effort—or undermine it by substituting active engagement. Its effects may therefore differ across student groups, raising concerns about widening educational gaps. In this paper, we provide the first experimental evidence on the causal effects of AI access on learning. We identify who benefits, who is left behind, and what can be done to mitigate emerging gaps.

We design and pre-register an experiment to answer these questions. Despite high interest in the topic due to the ongoing debate on whether and how to integrate generative AI into education, empirical evidence on its impact on learning remains scarce (Yan et al., 2024). The main challenge is that AI use is unobservable, making it difficult for field studies to restrict control groups from AI use. Few studies have aimed to establish causal effects (Bastani et al., 2024; Lehmann et al., 2024), and have been criticized due to features such as teacher quality that are hard to control in field settings (Tan and Rajaratnam, 2024). Our controlled lab experiment allows us to clearly identify the effects of AI on learning by ensuring full control over AI access—something that is difficult to achieve in field settings. This approach offers several important advantages. We can precisely control access to AI tools, ensuring a clear differentiation between treatment and control groups and avoiding spillovers. Our homogeneous setting guarantees that all participants learn about the same topic in the same way. Furthermore, our design enables a detailed analysis of distributional impacts across different student subpopulations, and our unique prompting data allows us to identify specific mechanisms through which AI use may help or hurt students. While the laboratory setting may be less realistic than the field, we design incentives that mirror real-course conditions and teach the material in a way that reflects how university students learn in practice. The experiment consists of three between subject treatment variations. Students in the lab are asked to go over a lecture and work on practice questions to learn about a topic that is new to them (Esperanto).[1] In the baseline treatment, students learn without access to AI but with access

[1]Esperanto is a constructed international language designed for easy learning and communication, with an estimated 100,000 to 2 million speakers worldwide. Its unfamiliarity to our sample and its simplicity make it ideal for our experiment, allowing students to learn a significant amount in a short time.

to Google Search, the status quo before the emergence of generative AI tools. In the AI-access treatment, students learn with access to AI, but with no formal guidance on how to use this tool except for some general instructions. In the AI-guided treatment, students learn with access to AI and are given guidance on how to use it best to promote their own learning.

Our experiment yields three key findings on the impact of AI tools on learning outcomes. First, contrary to widespread concerns among educators, we find no significant negative effects of AI use on learning, as measured by both exam performance and students' self-assessed learning. In fact we find that AI access *neither* hurts nor helps students on average. Second, our results reveal that this average zero effect masks significant heterogeneous effects across student subpopulations. While high-GPA women clearly benefit from AI-guided learning, male students show mixed responses. In particular, high-GPA men are less likely to score at the top or at the bottom, while low-GPA men experience increased variance in performance. Put differently, top women benefit from AI in their learning, while men -on average- remain unaffected. This means that AI risks exacerbating the pre-existing gender gap in learning outcomes as women -who have dominated men in educational outcomes for decades [2] - may end up doing even better. Third, we find that providing AI access changes the way students allocate their study time between chatting and active problem solving. Students in both AI treatment groups attempt significantly fewer practice questions within the same time limit as those in the control group who only have access to browsing. The reduced number of questions attempted when AI is available suggests a shift in attention away from active problem-solving toward chatting above and beyond the time spent browsing in the control treatment. This shift in time use may hinder skill development in the longer run.

Taken together, our findings suggest that AI can change not only educational outcomes but also how students learn in the first place. Women with High-GPA are expected to benefit the most from the AI-supported study, despite being the group with the lowest adoption rates in previous research (Carvajal et al., 2024). In contrast, mixed effects for male students raise concerns that AI could amplify existing educational inequalities. This is particularly relevant in light of the documented 'boy crisis' in education in OECD countries (Autor and Wasserman, 2013; Autor et al., 2023). Moreover, if students increasingly substitute AI interaction for active problem-solving, this change in study behavior may hinder the development of core skills and worsen learning outcomes over time.

[2] See for instance Voyer and Voyer (2014)

2 Experimental Design

We conducted the experiment in the CedEX Lab at the University of Nottigham in December 2024. Participants (N=572) are current undergraduate and graduate students from programs in all fields of study across the university.

The experiment was designed to mimic the stages of a typical university course, where students first learn about the new topic, then deepen and practice the content through practice questions, and finally are tested through an exam. To test how generative AI affects learning, we experimentally varied students' access to AI tools during the second stage of the experiment (completion of practice questions). In particular, participants were randomly assigned to one of three treatments, which are explained in more details below: T1: no access to ChatGPT (Control), T2: access to ChatGPT (AI-assisted), and T3: guided access to ChatGPT (AI-guided). All participants, independently of treatment status, learned about the same topic: Esperanto – a constructed international language designed for easy learning and communication. Its simplicity and unfamiliarity to the sample made it ideal for measuring how well students learn new material, rather than relying on their prior knowledge.[3]

Each session adhered to the same procedure:

1. **Stage 1**: Students learn Esperanto for 15 minutes using written learning materials provided by the researchers.

2. **Stage 2**: Students complete practice questions with access to different learning aids depending on treatment for about 20 minutes. This is the stage where the treatment variations take place.

3. **Stage 3**: Students take an exam testing their learning without access to any notes or learning aids.

4. **Stage 4**: Students take a post-study survey asking them about demographics as well as their learning experience during the session and attitudes and beliefs about AI.[4]

The entire experiment was conducted in Qualtrics. The learning materials and practice questions were presented across different pages, which allowed participants to move back and forth. More questions than possible to answer were included to accommodate

[3]This is a key design choice in our study because previous knowledge of a topic may discourage students to use AI since they can answer the questions themselves. We would then not be able to estimate the effect of AI on learning.

[4]Basic questions such as gender, experience with AI and languages spoken are asked at the beginning of the study, before stage 1.

different learning speeds. When the time for the lesson and practice questions was up, the screen moved automatically to a wait screen so that everybody would start the exam at the same time. There was no time limit on the exam but it was calibrated to last about 10 minutes on average. The total duration of each session was about one hour.

2.1 *T1: No access to ChatGPT*

T1 was designed to mirror learning before the advent of generative AI and to work as a baseline against which we can measure the impact of (guided) AI access on learning outcomes. As students had access to a web browser and online learning before the advancement of generative AI, we provided participants in this group free access to a web browser during stage 2. Websites providing access to AI chatbot sites, however, were blocked in T1. See instructions for students in Figure A5.

2.2 *T2: Access to ChatGPT*

T2 was designed to mirror learning when ChatGPT is available to students without clear guidelines, structure and training—a situation in which it is up to each student whether and how she or he wants to use AI in the learning process. This mirrors the situation which many universities face as they are grappling with the important task of adjusting their policies to the advent of AI.

Students accessed ChatGPT on their assigned computer browser with a already logged in premium ChatGPT account. The account was provided by the researchers and was cleared of any memory. The prompting history was archived and saved by researchers for analysis of prompting. See instructions for students in Figure A6.

2.3 *T3: Access to Guided ChatGPT*

This treatment is the exact same as T2, except that students now also receive guidance (written by the researchers) on how to use ChatGPT. The idea behind this treatment is to explore the potential of providing not just equal access but also equal training and encouragement in AI use for students. See instructions for students in Figure A7.

2.4 Incentives

We aimed to mimic the incentives provided in typical college courses in the lab. The incentives in most real courses around the world are to assign a smaller percentage of the final grade to assignments and practice questions and to have a larger weight on a final exam or midterms and a final exam. The idea of the assignments with lower percentages is to incentivize students to cover the materials and prepare them for the exam(s). In the same spirit, we offered incentives to reward correct responses and help students deepen their learning in the Esperanto lesson, while offering higher incentives for every correct answer in the exam. The structure of incentives is as follows:

- GBP 5 for completing the study
- GBP 7 for correctly answering at least 20 out of over 50 practice questions
- GBP 1 per correct answer in the test (maximum amount GBP 15)

The maximum payment they could earn is GBP 27.

From the beginning of the session students were informed that the most important part of the session (and where they can earn the most) is the exam. This is emphasized throughout the session. The intention is that they focus their efforts of trying to learn as much as possible form all the resources they have available, just as they would in a real-life course.

3 Empirical Strategy

3.1 Outcomes of interest

We focus on three primary measures of learning. First, our main variable of interest is exam performance, captured through both an objective and a subjective metric. The objective measure is the total number of correct answers on the final exam (ranging from 0 to 15). The subjective measure is students' self-assessed learning, recorded as an indicator equal to one if they slightly or strongly agreed with the statement "I feel like I learned a lot from solving practice questions" (on a four-point scale), and zero otherwise. For both measures, we investigate heterogeneous effects by background characteristics such as gender and GPA.

Second, we examine distributional outcomes by estimating the likelihood of students scoring in the top or bottom of the test score distribution. A top scorer is prespecified as having more than 10 correct answers, while a low scorer is defined as solving less than 5.

Finally, we assess students' exam preparedness by analyzing two measures: the number of practice questions solved correctly and the amount of time spent completing the final exam.

3.2 Econometric Specifications

The main specification regresses the outcomes presented above on indicators for the two AI treatments, keeping the control as the excluded indicator:

$$y_i = \alpha_0 + \alpha_1 \text{AI-assisted}_i + \alpha_2 \text{AI-guided}_i + \varepsilon_i \tag{1}$$

The main specification adds baseline covariates to control for different levels of AI adoption and students characteristics.[5]

We asses whether there are any gender differences in responses to the treatment using the specification:

$$\begin{aligned} y_i = \beta_0 + \beta_1 \text{AI-assisted}_i + \beta_2 \text{AI-guided}_i + \beta_3 \text{Female}_i + \beta_4 \text{AI-assisted}_i \times \text{Female}_i \\ + \beta_5 \text{AI-guided}_i \times \text{Female}_i + \varepsilon_i \end{aligned} \tag{2}$$

Where the coefficients β_1 and β_2 measure the effects of the treatments among men, β_3 measures any gender gaps in the control group, and β_4 and β_5 measure the differential effects of the treatments for women.

A similar specification will be presented replacing the Female variable with High GPA, a measure of academic skill equal to 1 if students are in first-class honours (GPA 70% or above) and 0 otherwise.

A final specification involves the interaction effects between treatments, gender and GPA. Following the previous literature, this specification may allow us to see differential effects by gender and GPA simultaneously. For example, if the treatments affect high-GPA female students differentially.

$$\begin{aligned} y_i = \gamma_0 + \gamma_1 \text{AI-assisted}_i + \gamma_2 \text{AI-guided}_i + \gamma_3 \text{Female}_i + \gamma_4 \text{High GPA}_i + \\ \gamma_5 \text{AI-assisted}_i \times \text{Female}_i + \gamma_6 \text{AI-assisted}_i \times \text{High GPA}_i + \gamma_7 \text{AI-guided}_i \times \text{Female}_i + \\ \gamma_8 \text{AI-guided}_i \times \text{High GPA}_i + \gamma_9 \text{AI-assisted}_i \times \text{Female}_i \times \text{High GPA}_i + \\ \gamma_{10} \text{AI-guided}_i \times \text{Female}_i \times \text{High GPA}_i + \varepsilon_i \end{aligned} \tag{3}$$

The triple interactions will measure whether top female students are differentially affected by the treatments. The double interactions with High GPA will measure if top male students are differentially affected by the treatments.

The results from these specifications are presented in table form in the appendix.

[5] The covariates include: Age 21 or younger, undergraduate student, indicators for field of study, use AI occasionally or all the time, has a paid subscription to an AI provider. In regressions without interactions with gender and high GPA, we add these variables as covariates.

4 Results

4.1 AI Has No Overall Negative Impact on Learning

We start by documenting the impacts of using AI on learning. This is a key outcome in the debate on whether to allow or forbid the use of generative AI tools at schools given that many educators and school administrators fear that AI will hinder student learning.

In Figure 1 we show two measures of learning: how well students perform on the exam, where they did not have access to external aids, and their self-assessment of learning. Figure 1; Panels (a) and (b) show the effects of AI guided and AI assisted on exam scores, respectively. The horizontal red line shows the average effect across all students, while the vertical lines show heterogeneous effects by gender and GPA relative to the control groups as pre-specified in our pre-analysis plan.

Overall, we do not find any effect of AI on exam scores. In AI guided, the mean treatment effect is -0.24 correct questions (p-value=0.188) relative to a base of 8.08 correct answers among students in the control group. The effect for AI assisted is also almost nil and not statistically significant at 0.08 additional correct answers (p-value=0.676).

There is also no effect on self-perceptions of learning, with a -4 percentage point (pp) effect (p-value=0.367) in AI guided and 1 pp effect (p-value=0.846) in AI assisted over a base of 81% of students in the control agreeing that they learned a lot in the session.

This overall null result masks interesting patterns across subgroups, especially in AI guided. Specifically, women who have access to AI guidance have significantly higher exam scores. However, this effect is not mirrored in their self-assessed learning: they are more pessimistic about their learning with AI than similar women in the control. The exam scores of men and low-GPA women do not seem to be affected on average by AI guidance. However, low-GPA women and in particular high-GPA men seem to think that they learn more when using AI. The patterns in AI assisted suggest that having access to the chatbot may hurt low-GPA men, and help low-GPA women. There are no salient patterns on self-assessed learning in AI assisted.

4.2 AI Widens Gaps: High-GPA Women Gain While Men Struggle

The above analysis of average scores suggests heterogeneity effects of having access to AI as a learning tool. We now explore this heterogeneity further by considering whether AI treatments affects who scores at the top or the bottom of the exam score distribution. For this we use the pre-specified indicators for top score (10 or more correct answers) and low score (5 or less correct answers). Figure 2 shows substantial heterogeneity in how AI access affects exam performance across subgroups. In Panel (a), we see that high-GPA score significantly higher on the exam when provided guided AI access. The opposite is

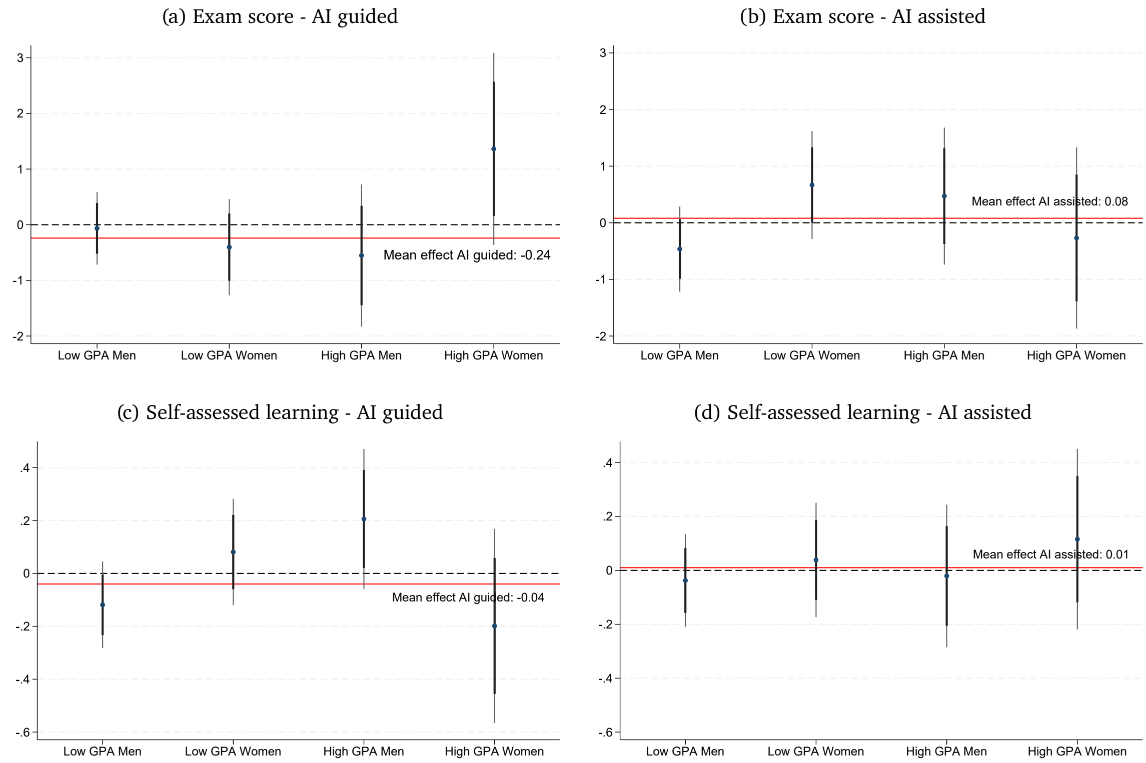

Figure 1: Treatment effects on exam score and learning

Notes: Panels (a) to (d) plot the coefficients from specification 3 for AI guided (left column) and AI assisted (right column). The outcome is stated in each panel title. The exam score is measured between 0 and 15 correct answers and the self-assessment of learning based on agreeing with the statement "I feel like I learned a lot from solving practice questions." In AI guided, the displayed coefficient on low GPA men is from the term γ_2AI-guided, on low GPA women from γ_7AI-guided × Female, on high GPA men from γ_8AI-guided × High GPA, and on high GPA women from γ_{10}AI-guided × Female × High GPA. Equivalent terms can be found for AI assisted in the regression. The red horizontal line represents the overall treatment effect. The point estimates for each of the subgroups include 95% confidence bars to visually assess whether the treatment effect is significantly different from zero, along with 83% confidence bars to visually assess differences between groups.

true for low-GPA women and high-GPA men who are less likely to obtain a top score in AI guided. In Panel (c) low- and high-GPA men again are differentially affected by AI guided, with the former group more likely while the latter less likely to get a low score. This evidence suggests that the variance in the distribution of scores among low-GPA men increases, while it becomes more concentrated for high-GPA men who are both less likely to get top and low scores. The effects of AI assisted (right panels) are not as conclusive.

4.3 AI Changes Study Behavior: Fewer Practice Questions Attempted

An important question is whether AI use distracts from traditional learning. For example, students may rely on the chatbot to answer questions without fully engaging with traditional course material, reducing the time spent actively studying. Conversely, they could use AI to engage with traditional material ask follow up questions etc, in which case they would spend more time than with traditional study methods than they normally would. Regardless, how students allocate their time between AI and conventional study tools is critical, as excessive use of AI may displace traditional learning, skill accumulation, and or leisure time.

We look at two dimensions of time use in Figure 3. The first in Panels (a) and (b) is time spent on solving the exam where no time limit was enforced, and access to AI was never provided. The second in Panels (c) and (d) is the number of practice questions attempted given the time limit of 37 minutes from the beginning of the session to the start of the exam with or without access to AI depending on treatment. Since the time is fixed in this outcome, we examine the number of questions students attempt to answer, which is another way of measuring time use. The time spent on the exam does not vary

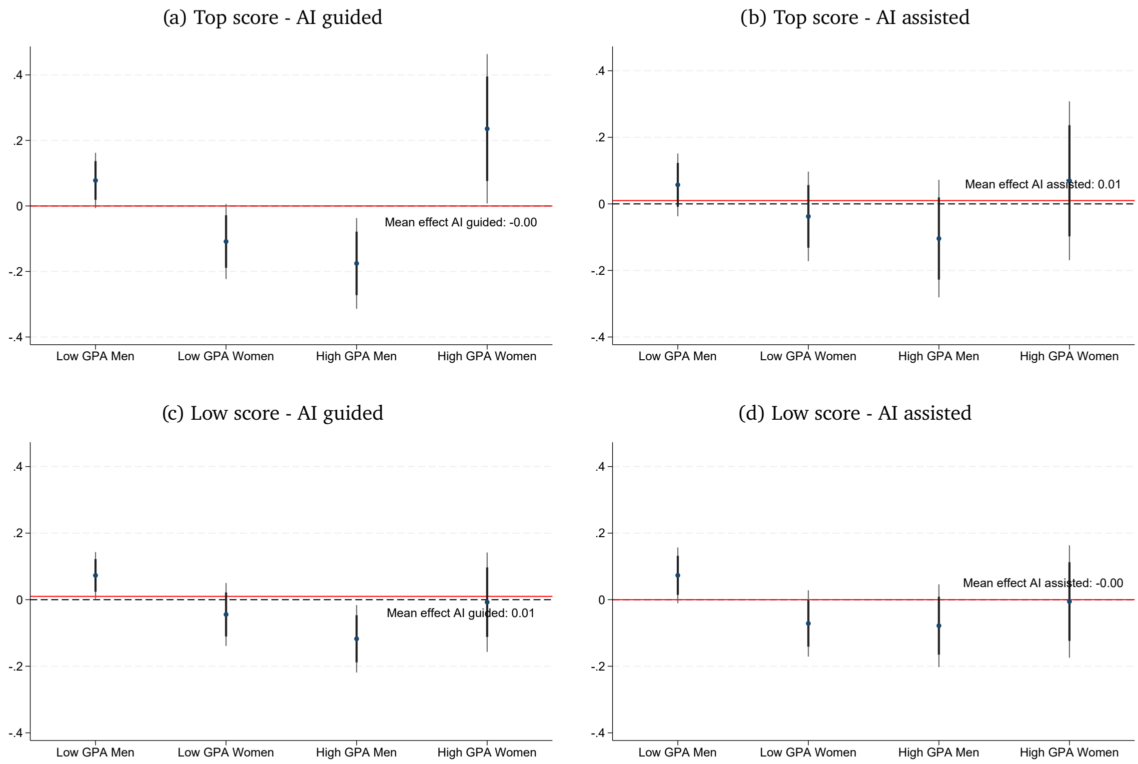

Figure 2: Treatment effects on obtaining a top or low score

Notes: Panels (a) to (d) plot the coefficients from specification 3 for AI guided (left column) and AI assisted (right column). The outcome is stated in each panel title. The top (low) scorer variables are indicators equal to 1 if the student scored 10 or more (5 or less) out of 15 possible points in the exam. In AI guided, the displayed coefficient on low GPA men is from the term γ_2AI-guided, on low GPA women from γ_7AI-guided $\times$ Female, on high GPA men from γ_8AI-guided $\times$ High GPA, and on high GPA women from γ_{10}AI-guided $\times$ Female $\times$ High GPA. Equivalent terms can be found for AI assisted in the regression. The red horizontal line represents the overall treatment effect. The point estimates for each of the subgroups include 95% confidence bars to visually assess whether the treatment effect is significantly different from zero, along with 83% confidence bars to visually assess differences between groups.

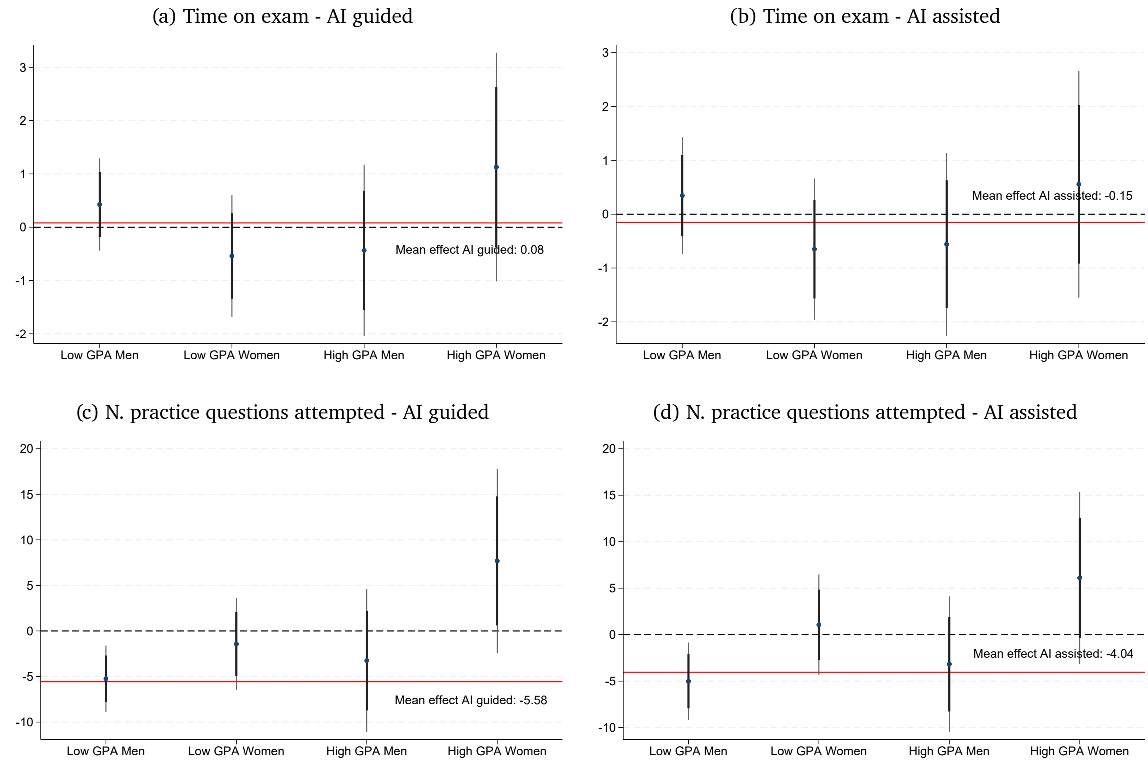

Figure 3: Treatment effects on time spent on exam and number of practice questions attempted

Notes: Panels (a) to (d) plot the coefficients from specification 3 for AI guided (left column) and AI assisted (right column). The outcome is stated in each panel title. In AI guided, the displayed coefficient on low GPA men is from the term γ_2AI-guided, on low GPA women from γ_7AI-guided × Female, on high GPA men from γ_8AI-guided × High GPA, and on high GPA women from γ_{10}AI-guided × Female × High GPA. Equivalent terms can be found for AI assisted in the regression. The red horizontal line represents the overall treatment effect. The point estimates for each of the subgroups include 95% confidence bars to visually assess whether the treatment effect is significantly different from zero, along with 83% confidence bars to visually assess differences between groups.

substantially across treatments or subgroups and is about 10.7 minutes on average. Since AI was not accesible for anyone at this stage it doesn't say much about how learning is affected by AI access. To answer this question, we turn to our second measure, number of practice questions attempted in different treatment arms. We see that having access to AI causally determines how many practice problems students attempt. On average, in the control group students attempt 34.5 questions. When students have access to AI, this is significantly lowered: by 5.58 questions attempted (p-value=0.000) in AI guided and by 4.04 questions (p-value=0.000) in AI assisted. Put differently, students attempt 16.2 % fewer practice questions when they study with guided AI access and 11.7 % fewer when they study with standard AI access, compared to those who study with only a browser. Both low- and high-GPA men are on the mark for the overall treatment effects, meaning that they attempt fewer questions than their control counterparts. We see again that women are the clear winners of providing AI access: in particular high-GPA women solve an equal amount of practice questions regardless of treatment, suggesting that having access to AI does not distract them from learning.[6]

5 Discussion

Our findings provide insights on the (heterogeneous) effects of AI on learning. The clear winners appear to be high-GPA women who perform better than men and than students in the control group in the AI guided treatment. Our evidence shows that a key reason for why this is the case is that they attempt as many practice questions as control students despite the overall and strong negative effect on number of attempted questions in the AI treatments. This suggests that they may have a more efficient use of AI. They self report that they use AI not just to answer the questions and that AI made them feel engaged with the materials. Paradoxically, high-ability women is the group that has been identified as having the lower adoption rates of AI (Carvajal et al., 2024). The analysis of the ChatGPT prompts will allow us to gain more insights as to why high-GPA women disproportionately benefit from AI use.

The potential losers in the AI revolution are the male and low GPA students. Not only were men attempting fewer practice questions but their performance was also below the control group. This can be seen more clearly in Figure 4, which plots the kernel densities of the exam scores for the whole sample and the main subgroups. The density for men in the AI treatments is well below and to the left of the control group. Students with low GPAs have a similar picture, while it is the opposite (AI has higher mass and to the right than the control distribution) for high-GPA students.

Based on these findings, it seems plausible that AI can exacerbate existing inequalities in education that have been summarized by the boy crisis, a term that was coined to characterize the phenomenon that boys are falling behind in education with detrimental consequences on their later-life outcomes (Autor and Wasserman, 2013; Autor et al., 2023). For example, in most OECD countries, a larger proportion of boys than girls do not attain the baseline level of proficiency in any of the core subjects, namely mathematics, reading and science, and drop out at higher rates than girls (OECD, 2015).

Overall, we see that some students may benefit from using AI in the learning process, in particular when there is guidance on how to use these tools. Our next steps will focus on analyzing the prompts of students in the AI treatments. This exercise will provide deeper insights into behavioral differences and help uncover why AI benefits some students more than others.

[6]Note however that students in both treatment groups solve a higher share of the practice questions that they attempted correctly, with no differences by student subgroups. Appendix Table A5 presents results from regressions using the share of correctly solved practice questions among all attempted practice questions as the outcome.

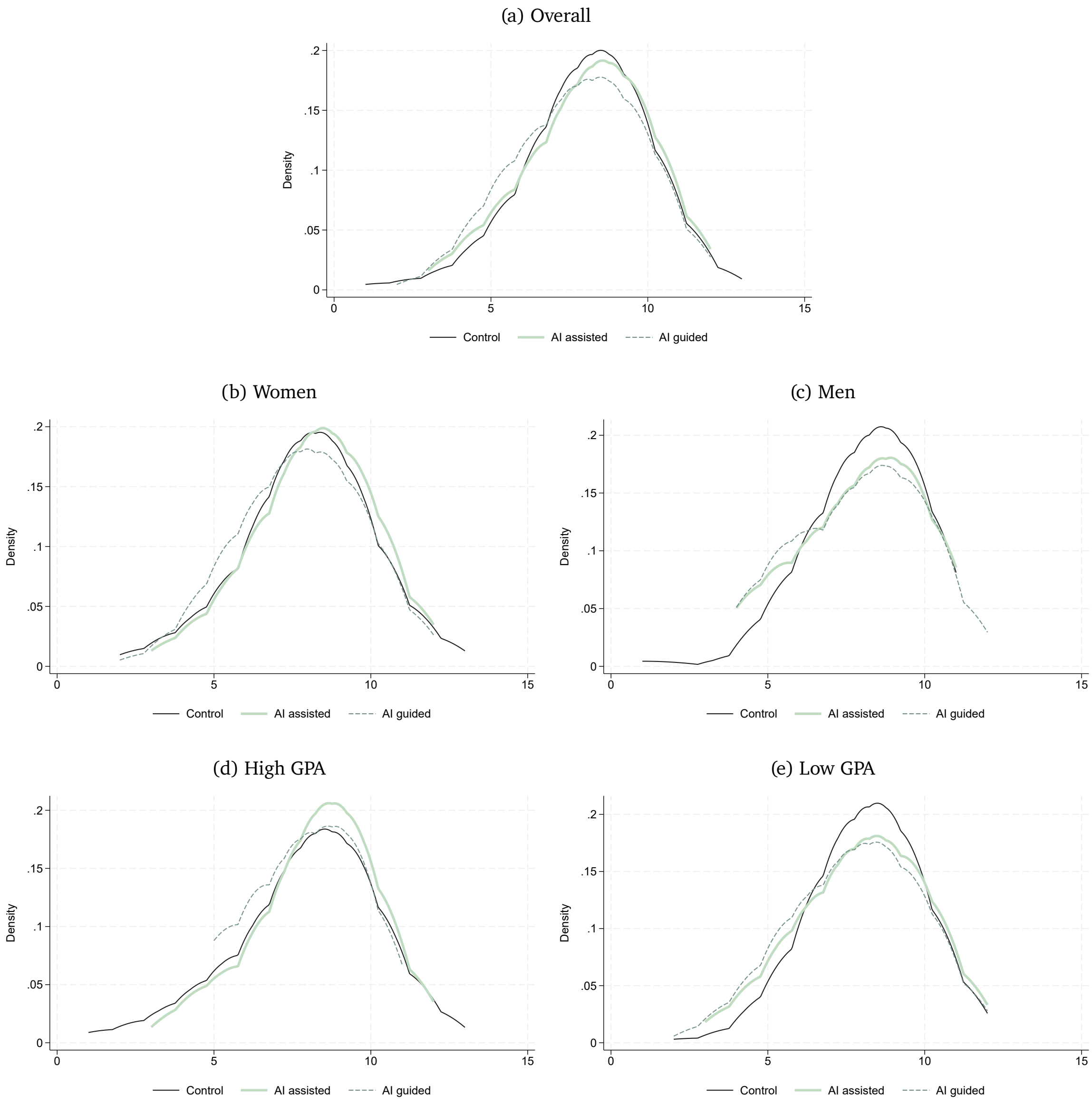

Figure 4: Kernel densities of exam scores

Notes: Panels (a) to (d) plot kernel densities for each of the treatment conditions by the pre-specified groups of interest. The kernel uses a smoothing bandwidth equal to 1.

References

Autor, David and Melanie Wasserman, "Wayward sons: The emerging gender gap in labor markets and education," *Third Way Report*, 2013, *20013.*

__ , David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman, "Males at the tails: How socioeconomic status shapes the gender gap," *The Economic Journal*, 2023, *133* (656), 3136–3152.

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Ozge Kabakcı, and Rei Mariman, "Generative ai can harm learning," *Available at SSRN*, 2024, *4895486.*

Carvajal, Daniel, Catalina Franco, and Siri Isaksson, "Will Artificial Intelligence Get in the Way of Achieving Gender Equality?," *NHH Dept. of Economics Discussion Paper*, 2024, (03).

Lehmann, Matthias, Philipp B Cornelius, and Fabian J Sting, "AI Meets the Classroom: When Does ChatGPT Harm Learning?," *arXiv preprint arXiv:2409.09047*, 2024.

OECD, "The ABC of Gender Equality in Education. Aptitude, Behavior, Confidence," 2015.

Tan, Steffi and Vaikunthan Rajaratnam, "Critique of Generative AI Can Harm Learning Study Design," *Available at SSRN 4898213*, 2024.

Voyer, Daniel and Susan D Voyer, "Gender differences in scholastic achievement: a meta-analysis.," *Psychological bulletin*, 2014, *140* (4), 1174.

Yan, Lixiang, Samuel Greiff, Ziwen Teuber, and Dragan Gašević, "Promises and challenges of generative artificial intelligence for human learning," *Nature Human Behaviour*, 2024, *8* (10), 1839–1850.

A Appendix Tables

Table A1: Balance of covariates across treatments

	(1) Control (T1)	(2) AI-assisted (T2)	(3) AI-guided (T3)	(4) Diff. (T1)-(T2)	(5) Diff. (T1)-(T3)
Female	0.618	0.613	0.602	0.004	0.016
	(0.487)	(0.488)	(0.491)	(0.049)	(0.049)
21 years old or younger	0.584	0.520	0.613	0.063	-0.029
	(0.494)	(0.501)	(0.488)	(0.049)	(0.049)
Undergraduate student	0.641	0.602	0.683	0.039	-0.042
	(0.481)	(0.491)	(0.466)	(0.048)	(0.047)
Arts	0.124	0.112	0.116	0.012	0.009
	(0.331)	(0.316)	(0.321)	(0.032)	(0.032)
Economics or Business School	0.201	0.224	0.241	-0.024	-0.040
	(0.402)	(0.418)	(0.429)	(0.041)	(0.041)
Engineering	0.163	0.163	0.131	-0.001	0.032
	(0.370)	(0.371)	(0.338)	(0.037)	(0.035)
Medicine and Health Sciences	0.163	0.153	0.146	0.010	0.017
	(0.370)	(0.361)	(0.354)	(0.036)	(0.036)
Social Sciences	0.153	0.199	0.176	-0.046	-0.023
	(0.361)	(0.400)	(0.382)	(0.038)	(0.037)
Other	0.196	0.148	0.191	0.048	0.005
	(0.398)	(0.356)	(0.394)	(0.037)	(0.039)
High GPA	0.368	0.429	0.226	-0.060	0.142^{***}
	(0.484)	(0.496)	(0.419)	(0.049)	(0.045)
Use AI occasionally or all the time	0.627	0.612	0.648	0.015	-0.021
	(0.485)	(0.488)	(0.479)	(0.048)	(0.048)
Paid AI subscription	0.057	0.056	0.090	0.001	-0.033
	(0.233)	(0.231)	(0.288)	(0.023)	(0.026)
Observations	209	196	199	405	408

Notes: Means and standard deviations (in parenthesis) based on questions asked at the beginning of the study. Differences and standard error of the difference (in parenthesis) in Columns 4 and 5. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table A2: Exam score and self-assessment of learning

	Test score				Self-assesment of learning			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-assisted	0.078	-0.224	-0.046	-0.464	0.004	-0.049	-0.014	-0.038
	(0.192)	(0.295)	(0.236)	(0.383)	(0.040)	(0.066)	(0.052)	(0.088)
AI-guided	-0.234	-0.211	-0.318	-0.066	-0.041	-0.069	-0.068	-0.119
	(0.189)	(0.293)	(0.217)	(0.332)	(0.041)	(0.067)	(0.048)	(0.083)
Female		-0.047		0.110		-0.025		-0.001
		(0.262)		(0.287)		(0.056)		(0.069)
AI-assisted × Female		0.498		0.667		0.087		0.039
		(0.386)		(0.485)		(0.083)		(0.108)
AI-guided × Female		-0.039		-0.403		0.045		0.081
		(0.390)		(0.440)		(0.085)		(0.102)
High GPA			0.006	0.288			-0.049	-0.012
			(0.290)	(0.441)			(0.059)	(0.093)
AI-assisted × High GPA			0.310	0.471			0.047	-0.021
			(0.397)	(0.616)			(0.083)	(0.135)
AI-guided × High GPA			0.258	-0.554			0.095	0.206
			(0.437)	(0.651)			(0.094)	(0.135)
Female × High GPA				-0.452				-0.059
				(0.593)				(0.120)
AI-assisted × Female × High GPA				-0.269				0.116
				(0.815)				(0.170)
AI-guided × Female × High GPA				1.362				-0.199
				(0.878)				(0.187)
Constant	7.448***	7.546***	7.514***	7.539***	0.828***	0.854***	0.843***	0.857***
	(0.311)	(0.322)	(0.311)	(0.318)	(0.067)	(0.069)	(0.068)	(0.075)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean women	7.968	7.968	7.968	7.968	0.808	0.808	0.808	0.808
Mean men	8.184	8.184	8.184	8.184	0.829	0.829	0.829	0.829
Observations	572	572	572	572	572	572	572	572

Notes: The outcomes are at the top of the columns. Estimates based on the specifications and heterogeneity reported in our pre-analysis plan. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table A3: Top and low score

	Top score				Low score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-assisted	0.009	0.016	0.034	0.057	-0.003	0.040	0.029	0.073*
	(0.028)	(0.041)	(0.035)	(0.048)	(0.020)	(0.029)	(0.024)	(0.043)
AI-guided	-0.002	0.026	0.010	0.078*	0.009	0.040	0.046**	0.073**
	(0.026)	(0.038)	(0.029)	(0.043)	(0.020)	(0.030)	(0.023)	(0.036)
Female		0.043		0.065*		0.029		0.015
		(0.035)		(0.038)		(0.025)		(0.020)
AI-assisted × Female		-0.011		-0.038		-0.070*		-0.071
		(0.056)		(0.068)		(0.039)		(0.051)
AI-guided × Female		-0.047		-0.109*		-0.050		-0.044
		(0.053)		(0.058)		(0.043)		(0.048)
High GPA			0.057	0.097			0.056*	0.031
			(0.042)	(0.062)			(0.032)	(0.035)
AI-assisted × High GPA			-0.061	-0.104			-0.081*	-0.078
			(0.059)	(0.090)			(0.043)	(0.063)
AI-guided × High GPA			-0.036	-0.175**			-0.123***	-0.117**
			(0.062)	(0.070)			(0.039)	(0.052)
Female × High GPA				-0.065				0.038
				(0.083)				(0.060)
AI-assisted × Female × High GPA				0.069				-0.006
				(0.121)				(0.086)
AI-guided × Female × High GPA				0.236**				-0.007
				(0.116)				(0.076)
Constant	0.022	0.012	0.010	-0.015	0.103***	0.079**	0.081**	0.064**
	(0.042)	(0.042)	(0.043)	(0.042)	(0.038)	(0.034)	(0.036)	(0.031)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean women	0.088	0.088	0.088	0.088	0.056	0.056	0.056	0.056
Mean men	0.053	0.053	0.053	0.053	0.013	0.013	0.013	0.013
Observations	572	572	572	572	572	572	572	572

Notes: The outcomes are at the top of the columns. Estimates based on the specifications and heterogeneity reported in our pre-analysis plan. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table A4: Time spent on exam and number of attempted questions

	Time spent on test (minutes)				Number of practice questions solved			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI-assisted	-0.132	0.162	-0.056	0.346	-4.171***	-6.255***	-4.337***	-5.022**
	(0.248)	(0.420)	(0.315)	(0.550)	(1.094)	(1.710)	(1.351)	(2.121)
AI-guided	0.149	0.225	0.090	0.426	-5.773***	-6.412***	-6.166***	-5.243***
	(0.243)	(0.371)	(0.290)	(0.441)	(1.125)	(1.751)	(1.305)	(1.852)
Female		0.479		0.904**		-1.865		0.085
		(0.377)		(0.455)		(1.538)		(1.691)
AI-assisted × Female		-0.484		-0.649		3.379		1.037
		(0.518)		(0.668)		(2.227)		(2.760)
AI-guided × Female		-0.124		-0.540		1.048		-1.469
		(0.494)		(0.583)		(2.333)		(2.580)
High GPA			0.279	1.004*			1.107	4.415
			(0.385)	(0.609)			(1.695)	(2.884)
AI-assisted × High GPA			-0.174	-0.560			0.460	-3.237
			(0.518)	(0.865)			(2.320)	(3.724)
AI-guided × High GPA			0.266	-0.435			1.396	-3.327
			(0.541)	(0.815)			(2.567)	(3.998)
Female × High GPA				-1.169				-5.357
				(0.783)				(3.484)
AI-assisted × Female × High GPA				0.555				6.197
				(1.072)				(4.731)
AI-guided × Female × High GPA				1.129				7.767
				(1.092)				(5.175)
Constant	10.112***	9.989***	10.109***	9.758***	30.538***	31.447***	30.718***	30.525***
	(0.435)	(0.472)	(0.446)	(0.510)	(1.991)	(2.025)	(1.946)	(2.072)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean women	10.819	10.819	10.819	10.819	33.857	33.857	33.857	33.857
Mean men	10.447	10.447	10.447	10.447	35.628	35.628	35.628	35.628
Observations	572	572	572	572	594	594	594	594

Notes: The outcomes are at the top of the columns. Estimates based on the specifications and heterogeneity reported in our pre-analysis plan. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table A5: Share of correctly solved practice questions

	Success rate in attempted practice questions			
	(1)	(2)	(3)	(4)
AI-assisted	0.055***	0.016	0.038*	-0.004
	(0.017)	(0.028)	(0.022)	(0.035)
AI-guided	0.062***	0.027	0.056***	0.021
	(0.017)	(0.026)	(0.020)	(0.031)
Female		-0.051**		-0.048*
		(0.023)		(0.029)
AI-assisted × Female		0.063*		0.064
		(0.035)		(0.045)
AI-guided × Female		0.057		0.056
		(0.035)		(0.041)
High GPA			0.008	0.008
			(0.023)	(0.037)
AI-assisted × High GPA			0.042	0.044
			(0.034)	(0.056)
AI-guided × High GPA			0.012	0.015
			(0.036)	(0.057)
Female × High GPA				-0.006
				(0.047)
AI-assisted × Female × High GPA				0.002
				(0.070)
AI-guided × Female × High GPA				-0.000
				(0.073)
Constant	0.685***	0.712***	0.692***	0.719***
	(0.028)	(0.030)	(0.028)	(0.031)
Controls	Yes	Yes	Yes	Yes
Mean women	0.678	0.678	0.678	0.678
Mean men	0.736	0.736	0.736	0.736
Observations	595	585	595	585

Notes: The outcomes are at the top of the columns. Estimates based on the specifications and heterogeneity reported in our pre-analysis plan. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

B Appendix Figures

You have been provided access to **Google Search and Google Translate** to assist you during the practice questions. These tools can help you look up information, find translations, and understand concepts from the material. You are free to decide how to use these tools to support your learning and answer the practice questions.

Here are some examples of how you might use the tools:

• You can find explanations or examples related to the grammar rules, such as:

- *"How do verbs change in Esperanto?"* or *"Esperanto plural nouns examples."*

• You can translate unfamiliar words or phrases between Esperanto and English.

• If you are curious about broader topics, you can search for additional context to enhance your understanding.

There are no specific instructions on how to use these tools, please use them as you normally would when you study for a topic in school. Some students may choose to use these tools frequently, while others may prefer relying on their notes or prior knowledge.

Reminders:

• The goal of this session is to engage with the material and complete the practice questions. Use the tools as you think is best to support your learning.

• You are encouraged to complete as many questions as possible within the time provided.

• There is no penalty for how much or how little you use the tools—you are free to decide how they fit into your learning process.

The goal is to learn as much as possible now. This is your chance to learn as much as possible with study aids to get the highest possible score in the final test where you will have no access to any external aids.

Figure A5: Instructions for control group (T1)

You have been given access to a **premium ChatGPT** account to assist you during the practice questions. This tool can help you look up information, find translations, and understand concepts from the material. You are free to decide how to use this tool to support your learning and answer the practice questions. Other generative AI chatbots are disabled and you cannot use your own premium account if you have one.

Please go now to the ChatGPT premium account that is open in your computer. **Type the following: Date_time_computer number.** For example, if the date is 28th November, 2024, the time of your session is 15:00, and your computer is #11, please write on the ChatGPT input box: 28112024_1500_11 and hit enter. You will then write all your interactions with ChatGPT in the same screen.

Here are some examples of how you might use ChatGPT:

• You can find explanations or examples related to the grammar rules, such as:

- "*How do verbs change in Esperanto?*" or "*Esperanto plural nouns examples.*"

• You can translate unfamiliar words or phrases between Esperanto and English.

• If you are curious about broader topics, you can search for additional context to enhance your understanding.

There are no specific instructions on how to use this tool, please use it as you normally would when you study for a topic in school. Some students may choose to use this tool frequently, while others may prefer relying on their notes or prior knowledge.

Reminders:

• The goal of this session is to engage with the material and complete the practice questions. Use the tool as you think is best to support your learning.

• You are encouraged to complete as many questions as possible within the time provided.

• There is no penalty for how much or how little you use the tool—you are free to decide how they fit into your learning process.

The goal is to learn as much as possible now. This is your chance to learn as much as possible with study aids to get the highest possible score in the final test where you will have no access to any external aids.

Figure A6: Instructions for AI-assisted group (T2)

You have been given access to a **premium ChatGPT** account to assist you during the practice questions. This tool can help you look up information, find translations, and understand concepts from the material. ChatGPT is a tool designed to help you with understanding the material and completing the practice questions. Other generative AI chatbots are disabled and you cannot use your own premium account if you have one.

Please go now to the ChatGPT premium account that is open in your computer. **Type the following: Date_time_computer number.** For example, if the date is 28th November, 2024, the time of your session is 15:00, and your computer is #11, please write on the ChatGPT input box: 28112024_1500_11 and hit enter. You will then write all your interactions with ChatGPT in the same screen.

To make the most of ChatGPT as a learning tool, follow these steps:

1. Explain your thought process or question. Before typing, **describe what you're trying to figure out or learn**. For example: *"I'm learning about Esperanto nouns and their endings. Can you explain how plural nouns work?"*

2. Ask for step-by-step guidance. Rather than asking for direct answers, request explanations or steps. For example: *"Can you explain step-by-step how to form a past-tense sentence in Esperanto?"*

3. Break questions into smaller parts. **Focus on one aspect at a time.** For instance:
"What is the ending for adjectives in Esperanto?"
Follow up with: *"Do adjectives also change in the plural form?"*

4. Engage in follow-up questions. Clarify and expand on ChatGPT's responses. For example: *"Can you give me more examples of how to use the word 'bela' (beautiful) in different contexts?"*

5. Practice actively. Write your own sentences and **ask ChatGPT to check them or provide feedback**: *"I wrote: 'La granda kato dormas.' Is this sentence correct?"*

Reminders:
- The goal is to use ChatGPT not just to find answers but to deepen your understanding and learn how to approach the material thoughtfully.
- You are encouraged to ask questions, experiment with your understanding, and clarify concepts during the session.
- Take your time to engage fully with the tool and explore how it can help you learn.

The goal is to learn as much as possible now. This is your chance to learn as much as possible with study aids to get the highest possible score in the final test where you will have no access to any external aids.

Figure A7: Instructions for AI-guided group (T3)