Table 4: Effect of GenAI on total debating points by sub-category and student ability

	Clarity		Credibility		Rhetoric		Refutation		Superiority of arguments	
	Low	High	Low	High	Low	High	Low	High	Low	High
Treat	0.238**	-0.197	-0.134	1.756*	0.081	1.736**	0.126	1.858**	0.090	1.430**
	[0.007]	[0.496]	[0.527]	[0.045]	[0.412]	[0.010]	[0.253]	[0.006]	[0.356]	[0.006]
Mean dep. var.	6.12	6.02	5.12	5.28	6.41	6.22	5.82	5.50	5.94	5.67
SD dep. var.	0.73	1.07	1.73	1.62	0.81	0.84	0.93	1.14	0.81	1.04
R^2	0.30	0.61	0.20	0.62	0.31	0.60	0.23	0.64	0.22	0.61
Obs.	304	60	304	60	304	60	304	60	304	60

Significance levels are indicated by $^{+} < 0.1$, * < .05, ** < .01, *** < .001. The table shows coefficients from regressions of equation 1, where the outcome is the total debating points of individual i in debate round r for each rubric sub-category (averaged across the judges evaluating the same debate), estimated separately for low-ability (non–merit-scholarship) and high-ability (merit-scholarship) students. All specifications include the same controls as Table 2. Because the high-ability subsample contains only 20 individuals (clusters), inference is based on the wild cluster bootstrap for both panels, with 9,999 replications, Rademacher weights, the null imposed, and clustering at the individual level. Values in brackets are wild-cluster-bootstrap p-values; significance stars derive from the same test. Each individual is observed in 2–3 debate rounds. The sample is 141 of 142 randomized individuals; one did not complete the end-line survey.

A Online Appendix

A.1 Consent Form

The ESADE Debate Challenge is a debate competition in which participants are provided with various tools and resources to prepare for debates. Ten prizes of €100 will be awarded in total, five in each category, depending on the tools with which the participants compete. In the assessments, students compete only against students in their own category.

The purpose of this study is to test the impact of different technological tools in democratic debates. To this end, the information (audios) collected during this event will be analyzed anonymously and scientifically by the ESADE researchers. Therefore, by agreeing to participate in the ESADE Debate Challenge, you also agree to participate in a scientific study. The characteristics of the study are explained in more detail below.

This project has been approved by the ESADE Research Ethics Committee (CUHSR protocol number: 011/2023).

You must be at least 18 years of age to participate in this study or provide informed consent from your parent/guardian.

If you agree to participate in this study, you will be asked to do the following:

- Complete an online questionnaire before the in-person debate.
- Participate in a debate competition that includes four debates and complete three very brief surveys.
- The participants themselves will record the debate on their cell phones and send the recording to an ESADE phone number where all recordings will be centralized. The files will be identified only by a code and will not be used or made available for any purpose other than the research project. The files will be destroyed at the end of the study.
- Participation in this study will take a total of 180 minutes of your time, but may be slightly extended due to logistical difficulties.
- There are no known risks beyond everyday life associated with your participation in this study
- All participants will receive a certificate from EsadeEcPol for their participation. Ten prizes of 100 euros will be awarded to the winners (5 in each of the two categories).
- Participation in this study is voluntary and you can withdraw from it at any time. You will not receive any direct benefit from the study.
- If you have questions about your rights and welfare as a volunteer participant in the study, please contact the Esade Research Ethics Committee at ethics@esade.edu.

- The confidentiality of your research records will be strictly maintained by ensuring that all data will be kept secure and that only the principal investigator and the research team will have access to these data. This means that no one else will have access to your data at any time during or after the study.

Basic information about the processing of your personal data

- Data controller: The data controller is the ESADE Foundation.
- Contact: lopd@esade.es
- Purpose: Consent to participate in research studies
- Legal basis: Consent, legitimate interest in the research studies and compliance with legal obligations.
- Addressee: Esade EcPol, Esade Decision Lab - Research Office, Principal Investigator(s)
- Data management: Data will be deleted when it is no longer necessary to fulfill the purpose for which it was collected. The most relevant information will be retained permanently. The criteria for retention or deletion will be based on public records regulations or result from the performance of public duties.
- Rights: you have the right to request from the controller information about your personal data and its rectification or erasure, or to restrict processing, or to object to processing, as well as the right to data portability, the right to withdraw your consent at any time, and the right to lodge a complaint with a supervisory authority.

By checking the box below, you agree to participate in the study and acknowledge that you have read, understand, accept and will comply with the above instructions and conditions.

- I agree

A.2 Baseline Survey

1. Full name
2. Gender (Male, Female)
3. Date of birth (Drop down)
4. Do you feel comfortable talking about complex topics in English?
   - ☐ Yes, I have no issues
   - ☐ Yes, quite comfortable
   - ☐ No, but I can hold my own
   - ☐ No, I find it very difficult
5. Do you feel comfortable talking about complex topics in Spanish?
   - ☐ Yes, I have no issues
   - ☐ Yes, quite comfortable
   - ☐ No, but I can hold my own
   - ☐ No, I find it very difficult
6. Degree you are currently pursuing:
   - ☐ Law and International Relations
   - ☐ Law
   - ☐ Economics
   - ☐ Business Administration
   - ☐ Other
7. University where you are pursuing these studies:
   - ☐ ESADE
   - ☐ Other
8. In your last high school year, on a scale of 0 to 10, where 0 is the lowest grade and 10 is the highest grade, what was approximately your average grade at the end of the year?
   - ☐ (0-2)
   - ☐ (2-4)
   - ☐ (4-6)
   - ☐ (6-8)
   - ☐ (8-10)
9. Are you currently receiving any type of scholarship?
   - ☐ Yes, an academic excellence scholarship
   - ☐ Yes, other income-related scholarship

☐ No

10. What is the highest level of education completed by your father?

    ☐ Primary school
    ☐ Secondary education
    ☐ Professional training
    ☐ University degree
    ☐ Master's degree
    ☐ PhD

11. What is the highest level of education completed by your mother?

    ☐ Primary school
    ☐ Secondary education
    ☐ Professional training
    ☐ University degree
    ☐ Master's degree
    ☐ PhD

12. On a scale from 0 to 100, with 100 being "a lot" and 0 being "not at all", how much do you enjoy speaking in public?

13. Have you participated in a debate competition before?

    ☐ Yes, several times
    ☐ Yes, once
    ☐ No, never

14. Have you ever received a prize in a debate competition?

    ☐ Yes, several times
    ☐ Yes, once
    ☐ No, never

15. On a typical day, approximately how much time do you spend watching, reading or listening to news about politics and current affairs? Please answer in hours and minutes. For example, if you spend one hour and twenty minutes, you would enter 01 under "HOURS" and 20 under "MINUTES".

16. In politics, we sometimes refer to the "left" and "right". Where would you place yourself on this scale? 0 means "left" and 10 means "right".

17. On a scale where 0 means you have very unfavourable feelings and 100 means you have very favourable feelings towards people who hold political views that are opposite to yours, where do you position yourself? A value of 50 means that your feelings are neither favourable nor unfavourable.

18. Which of the following software do you know?

    ☐ Google Drive
    ☐ Tableau
    ☐ ChatGPT
    ☐ Overleaf
    ☐ Jasper
    ☐ Grammarly

19. Which of the following software have you used recently?

    ☐ Google Drive
    ☐ Tableau
    ☐ ChatGPT
    ☐ Overleaf
    ☐ Jasper
    ☐ Grammarly

20. Price controls: Please indicate your level of agreement with the following statements: [0=Strongly disagree 100=Strongly agree].

    (a) A rent price control system should be implemented in all medium and large cities (0-100).
    (b) Price control systems are a bad idea (0-100)
    (c) Implementing a rent price control system in all medium and large cities would have positive consequences (0-100)
    (d) Do you agree or disagree with the implementation of a rent control system in all medium and large cities and neighbourhoods? [0=Strongly disagree 100=Strongly agree].
    (e) If there was a referendum tomorrow on implementing a system of rent price controls in all medium and large neighbourhoods and cities, how likely is it that you will vote in favour? [0=would NOT vote in favour with 100% certainty; 100 =would vote in favour with 100% certainty].

21. Central Bank Political Control: Please indicate your level of agreement with the following statements: [0=Strongly disagree 100=Strongly agree].

    (a) EU governments should regain political control of the ECB in order to be able to finance themselves on more advantageous terms (0-100)
    (b) The ECB's monetary policy should be subordinated to the fiscal needs of the member states (0-100)
    (c) In a situation where an EU member state is forced to make cuts, it is always better for the ECB to offer an unconditional bailout to that state (0-100)
    (d) Do you agree or disagree with EU member state governments regaining political control over the European Central Bank and thus, over monetary policy? [0=Strongly disagree 100=Strongly agree]

(e) If there was a referendum tomorrow on regaining political control over the direction of the European Central Bank and its monetary policy, how likely is it that you will vote in favor? [0=would NOT vote in favor with 100% certainty; 100 =would vote in favor with 100% certainty].

22. Retirement age and youth employment: Please indicate your level of agreement with the following statements: [0=Strongly disagree 100=Strongly agree].

    (a) Lowering the retirement age is a good measure to increase youth employment (0-100)

    (b) If there are 100 jobs in society, it is the government's responsibility to ensure that older people retire earlier in order to free up jobs for younger workers (0-100)

    (c) Lowering the retirement age from 67 to 60 would significantly reduce youth unemployment without negative effects (0-100)

    (d) Do you support or oppose the government's intervention to distribute jobs more equitably among different generations in society? [0=Strongly disagree 100=Strongly agree]

    (e) If there was a referendum tomorrow to lower the retirement age in order to increase youth employment, how likely is it that you will vote in favour? [0=would NOT vote in favour with 100% certainty; 100 =would vote in favour with 100% certainty].

23. Job guarantee: Please indicate your level of agreement with the following statements: [0=Strongly disagree 100=Strongly agree]

    (a) If a country has unemployed workers, the state should offer them a job through a job guarantee program (0-100)

    (b) A job guarantee program would lower the unemployment rate without negatively affecting other economic indicators (0-100)

    (c) Guaranteeing employment for all people of working age should be a recognised right, regardless of the public expenditure involved (0-100)

    (d) Are you in favour or against the government passing a law guaranteeing public employment for all unemployed people? [0=Strongly disagree 100=Strongly agree]

    (e) If there were a referendum on such a law tomorrow, how likely is it that you will vote in favour? [0=would NOT vote in favour with 100% certainty; 100 =would vote in favour with 100% certainty].

24. Taxes and tax collection: Please indicate your level of agreement with the following statements: [0=Strongly disagree 100=Strongly agree]

    (a) Reductions in taxes on labor, such as the personal income tax, cause people to work more and ultimately increase tax revenues. (0-100)

    (b) A tax increase leads to a taxpayer response in the form of lower consumption and employment, which ultimately reduces tax revenues. (0-100)

(c) When there are tax cuts in a country/region, workers from other regions move to where taxes are lower, which helps to increase final revenues (0-100)

(d) Are you in favor or against the government cutting taxes such as personal income tax or VAT? [0=Strongly disagree 100=Strongly agree]

(e) If there were a referendum tomorrow on lowering the general VAT rate from 21% to 10% and cutting income tax in half, how likely is it that you will vote in favor? [0=would NOT vote in favor with 100% certainty; 100 =would vote in favor with 100% certainty].

25. Determinants of social mobility: Please indicate your level of agreement with the following statements: [0=Strongly disagree 100=Strongly agree]

(a) The economic situation of the parents in childhood is not relevant for the future of a person. (0-100)

(b) A person's effort is the main determinant of his or her success in life. (0-100)

(c) The effort of a person from a poor background is rewarded in the same way as those of a person from a wealthy family. (0-100)

(d) Are you for or against the government redistributing income from the rich to the poor in order to reduce inequality of opportunity? [0=Strongly disagree 100=Strongly agree]

(e) If there were a referendum tomorrow to eliminate all wealth and inheritance taxes, how likely is it that you will vote in favour? [0=would NOT vote in favour with 100% certainty; 100 =would vote in favour with 100% certainty].

26. School repetition: Please indicate your level of agreement with the following statements: [0=Strongly disagree 100=Strongly agree]

(a) Repeating a year is an effective measure to improve the level of learning in the vast majority of cases where it is applied (0-100)

(b) Repeating a year does not increase the probability that a student will drop out of the educational system early (0-100)

(c) In terms of cost-benefit, repeating a year is a superior educational policy to tutoring in small groups or other reinforcement programs (0-100)

(d) Do you agree or disagree with the government passing a law that severely limits the cases in which a student can be required to repeat a year? [0=Strongly disagree 100=Strongly agree]

(e) If there was a referendum tomorrow to limit the cases in which a student can be required to repeat a year, how likely is it that you would vote in favor of it? [0=would NOT vote in favor with 100% certainty; 100 =would vote in favor with 100% certainty].

27. Trade policy: Please indicate your level of agreement with the following statements: [0=Strongly disagree 100=Strongly agree]

(a) The interests of national industry should be considered before opening trade with any country (0-100)

(b) Even if a product is much cheaper abroad, it should not be imported if it would result in job losses in national industry (0-100)

(c) It is preferable, from an economic efficiency standpoint, for consumers to pay higher prices so that national industry does not lose jobs (0-100)

(d) Do you agree or disagree with the government passing a law establishing a tariff on products imported from other countries to make domestic products more attractive? [0=Strongly disagree 100=Strongly agree]

(e) If there was a referendum tomorrow to limit imports from third countries, how likely is it that you would vote in favor of it? [0=would NOT vote in favor with 100% certainty; 100 =would vote in favor with 100% certainty].

A.3 Debate challenge regulation (sent to ESADE students)

The following document establishes the rules for the ESADE Debate Challenge, acceptance of which is a prerequisite for registration and participation in the competition. In addition, all participants must complete an Initial Questionnaire which includes a declaration of consent that the data collected during the event will be used for academic purposes.

A.3.1 Article 1. Eligibility

The competition is open to students from all colleges and universities and to participants in the Pre-University Debate League. Contestants may be asked to show proof of enrolment in university or in the Pre-University Debate League. In addition, minors under the age of 18 must provide a signed consent form from their parent or guardian.

A.3.2 Article 2. Required materials

All students must bring their laptop and a cell phone capable of audio recording, as well as the appropriate chargers for both devices, as they are essential for participation in the competition.

A.3.3 Article 3. Topics to be debated

There are eight different topics, each with its corresponding debate question. Each participant will have to defend the position assigned to him or her at random, either for or against, regardless of his or her personal opinions on the subject. The selected topics deal with current issues in the field of public policy. On the day of the debate, four of these eight topics will be randomly selected and discussed by the participants.

A.3.4 Article 4. Date, language and venue

The confrontation rounds will be held on March 20 and 24 in the auditorium of ESADE Creapolis in Sant Cugat del Vallès. On March 20, the debates will be held in English, while on March 24 they will be held in Spanish. The schedule for Monday, March 20, is from 3:00 p.m. to 6:30 p.m., and for Friday, March 24, from 3:30 p.m. to 7:45 p.m. All participants should arrive at the Creapolis Building auditorium 15 minutes before the designated time to register and begin on time.

A.3.5 Article 5. Structure and sequence of the individual discussion rounds

Each participant competes individually. The debates take place in the auditorium and are organised as direct confrontation between the speakers. After being briefed on the topic to be debated and the assigned position, 20 minutes are allowed for preparation. The debate lasts 10 minutes and proceeds as follows:

- First, the speaker with the position in favour begins the debate with a 3-minute presentation.
- Next, the speaker representing the opposing side gives their first 3-minute presentation.
- Then, the speaker in favour has 2 minutes to respond to their opponent's arguments.
- To conclude the debate, the speaker of the opposing side has another 2 minutes to address the counter-arguments of his opponent.

A.3.6 Article 6. Preparation tools and resources

This tournament is part of a research study in which participants will be randomly divided into two different groups. Each group will have access to different sources and tools to prepare for their respective debates, which may initially lead to some inequities. However, all participants will be evaluated based on the resources assigned to their respective groups, and prizes will be awarded based on individual scores, taking into account the tools available to each participant.

Speakers will have the option of reading their speech directly from a piece of paper or a screen if they deem it appropriate. They may also use brief notes to remember the main points of their argument.

A.3.7 Article 7. Communication between groups

From the moment the participants have been divided into two groups, communication between the members of the two groups is prohibited, except during the debate. Failure to comply with this rule may result in the participant's exclusion from the debate tournament.

A.3.8 Article 8. Recording of the debates

The debates will be recorded using the cell phone of one of the participants for later analysis. Below are the instructions for recording and sending the files:

- Before you begin recording, enable airplane mode on the mobile device being used for recording. This will avoid interruptions from incoming calls.
- Use an audio recording application pre-installed on the phone, such as "Voice Memos" on iPhone or "Recorder" on Android.
- Before the start of the debate, participants should state their identification number (assigned to them upon arrival), the question to be debated, and the position defended to facilitate file identification.

- The cell phone will be passed between the debaters as if it were a microphone to achieve better sound quality, considering that there will be more people in the room.
- Speak in a moderate tone of voice to ensure proper recording.
- It is recommended that participants bring a charged cell phone battery and charger if needed.
- At the end of the discussion and before leaving the table, the audio file will be sent via WhatsApp to +34 645 155 884.
- Before leaving the table, please notify someone from the organisation to verify that the audio file was received correctly.
- Once the recording has been sent via WhatsApp, the file will be uploaded to a OneDrive folder within 24 hours. The link to the corresponding folder will be provided at the end of the event.

A.3.9 Article 9. Evaluation of the debates

The recordings of the debates will be judged by a panel of experts in the field of debates. Each debate will be evaluated by three different judges. The evaluation will focus primarily on the substantive aspects of the debate, although the rhetorical and oral skills of the participants will be also considered. In this way, the overall performance of each speaker will be evaluated. The jury will decide which of the speakers are the winners. The evaluation will take place over a period of three to four weeks after the debate.

A.3.10 Article 10. Evaluation Criteria

Participants will be evaluated in each of the following categories [10] :

1. Ability to respond to the question posed.
2. Coherence between thesis and argument.
3. Clarity of presentation: the arguments are easily recognisable.
4. Correctness of argument: it is not limited to the presentation of evidence.
5. Credible and adequately presented evidence.
6. Variety of evidence.
7. Appropriate and persuasive use of language.
8. Ability of the arguments to respond to the opponent.
9. Superiority of arguments compared to those of the opponent.
10. Ability of arguments to integrate counterarguments.

[10]These ten cathegories where consolidated into a five point rubric for convenience of the judges' evaluation, following the recommendation of debate judges advising in the process.

A.3.11 Article 11. Prizes and recognition

Cash prizes in the form of Amazon gift cards will be awarded to the participants who achieve the ten highest scores in the competition. Since the debaters have different resources to prepare their arguments, the prizes will be awarded considering the participants in equal categories. A total of 10 prizes of 100 euros each will be awarded. In addition, the remaining participants will receive a participation diploma awarded by EsadeEcPol, ESADE's Center for Economic Policy, in recognition of their efforts and commitment to the competition.

A.3.12 Article 12. Changes in rules and regulations

The Organizing Committee reserves the right to make changes to the rules and regulations at any time and without prior notice.

A.4 Topics debated

The following topics were debated during the debate competitions:

1. Rent Controls: ¿Should the state set housing prices?
2. Job Guarantee: Should the State guarantee the full employment of the working-age population by directly providing jobs to the unemployed?
3. Central Bank Political Control: Should EU governments regain political control of the ECB in order to finance themselves on more advantageous terms and avoid austerity?
4. Retirement age and young employment: Would lowering the retirement age help young people to find work?
5. Taxes and tax collection: Does lowering taxes help improve tax collection?
6. Determinants of social mobility: Is the family socio-economic background a strong determinant of people's job opportunities in life?
7. School repetition: Is repeating a course an effective way to improve the level of learning?
8. Trade policy: Should the government punish or even forbid products from third countries that may pose a threat to the national industry?

A.5 Rubric to score debates

Please, assign 0 to 10 points per category to each participant. You may conclude that both debaters deserve 10 or 0 points in the same category, at your discretion. Remember: your mission is to assess the validity of the argument construction mainly, not so much the form that it acquires. Focus on examining the substance and content being debated. "Total score" should reflect the sum of the previous five scores and should be higher for the winner of the debate (there cannot be a tie).

Table A1: Rubric for debate scoring

Dimension	Evaluation	Points given to position "In favor"	Points given to position "Against"
1.	Clarity, correctness and validity of the defended position.		
2.	Credible and well-presented evidence.		
3.	Formal quality and rhetoric of the participant (not the content). Convincing use of language.		
4.	Ability to refute the rival's position		
5.	Superiority of the arguments to those of the rival.		
	Total score:		

A.6 Deviations from pre-registration

The experiment was registered in the AEA RCT Registry in March 2023. I registered four research questions. In this paper, I only address the first two questions: (a) whether GenAI improves the debating skills of the participants; (b) whether it contributes to reducing the inequality in the debating skills. The main outcome variables I use are the same that were pre-registered: (a) debating points and (b) winning a debate. However, while I tried to access administrative data from students to analyze possible longer-term effects on grades and other relevant variables, I received a very low response rate from students when asked for permission to access their data. A relevant deviation from the preanalysis plan is that in order to measure the heterogeneous effects of AI on high-skilled and low-skilled students, I not only use the performance of participants in the first round of debates - top and bottom (50%), but I also use the merit scholarship variable. This was a question I added to the baseline survey that I think works as a very good complementary measure of innate ability because it is not directly connected with the experiment and does not depend on the performance in a single debate, which could be noisy.

A.7 Additional figures

Figure A1: Timeline of the experiment

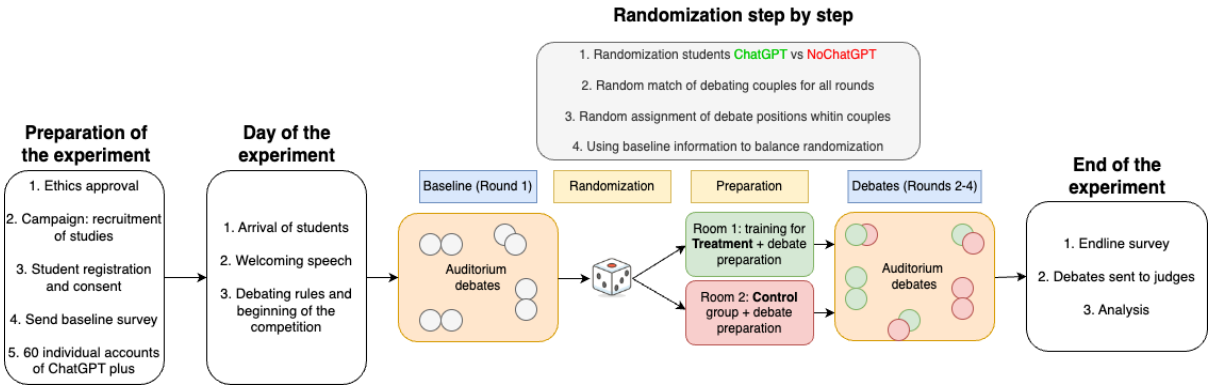
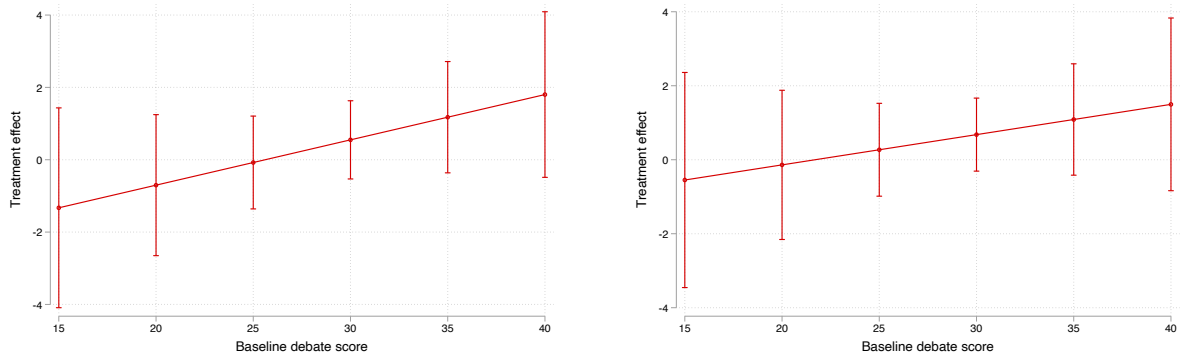


Figure A2: Heterogeneous Treatment Effects by Continuous Baseline Debate Score

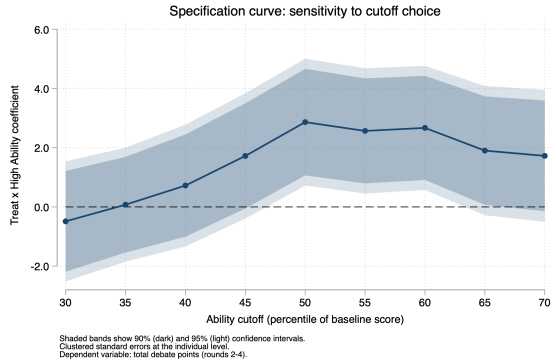
(a) Without controls *(b) With controls*



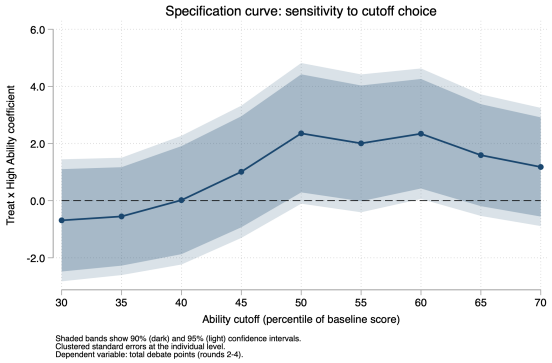
Notes: The figure plots the marginal effect of GenAI access on total debate points as a function of the baseline debate score, evaluated at baseline-score $\in \{15, 20, 25, 30, 35, 40\}$. Margins are computed from the following OLS specification estimated on post-treatment rounds (round > 1), with standard errors clustered at the individual level: $Y_{ird} = \alpha_d + \alpha_r + \beta\,\text{Treat}_i + \gamma\,S_i + \delta\,(\text{Treat}_i \times S_i) + \lambda X_i' + \varepsilon_{ird}$, where Y_{ird} is total debate points for individual i in round r on debate day d; S_i is the continuous baseline debate score; α_d and α_r are debate-day and debate-round fixed effects respectively; and X_i is the vector of individual controls included in column (2) of Table 2 (age, gender, parental education, economics background, prior debating experience, prior GenAI use, scholarship status, language comfort, and high-school GPA). Panel (a) omits X_i; panel (b) includes it. Spikes represent 90% confidence intervals.

Figure A3: Specification Curve: Sensitivity of the Treat × High Ability Interaction to Cutoff Choice

(a) Without controls

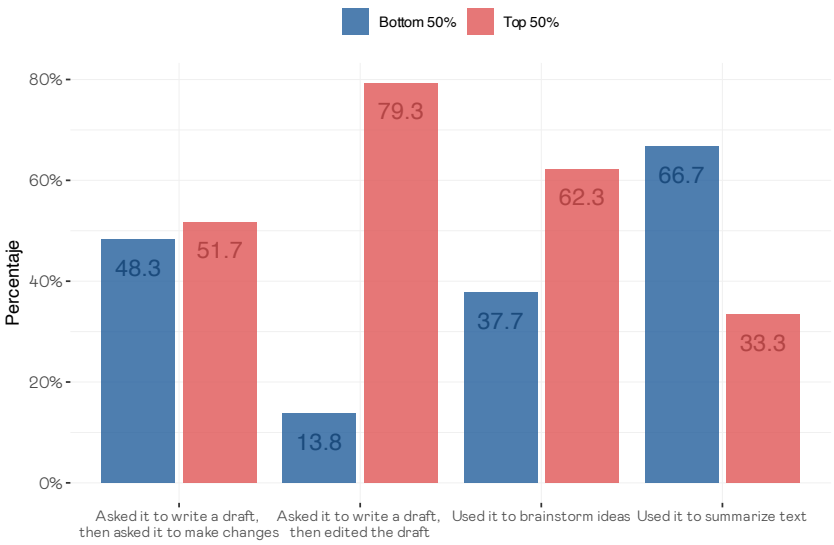


(b) With controls



Notes: Each point shows the estimated coefficient on the Treat × High Ability interaction term from equation 2, where "High Ability" is defined as scoring at or above a given percentile of the baseline debate score distribution. The percentile cutoff varies from the 30th to the 70th in steps of five. For each cutoff, the binary ability indicator is constructed using `xtile` on round-1 debate scores. The regression is estimated on post-treatment rounds round > 1) with debate-day and debate-round fixed effects and baseline debate score as an additional control; standard errors are clustered at the individual level. Panel (a) omits individual controls; panel (b) includes the full control vector described in the notes to Table 2. Shaded areas represent 90% (dark) and 95% (light) confidence intervals.

Figure A4: How the treatment group used GenAI, by student ability



A.8 Additional tables

Table A2: Balance by baseline performance

	(1) Top 50%	(2) Bottom 50%	(3) Difference (1)-(2)	(4) p-value Col. (3)
Age	20.19	20.45	-0.26	0.32
Female	0.44	0.49	-0.04	0.60
Father holds Master or higher degree	0.50	0.38	0.12	0.14
Economics background	0.50	0.50	0.00	1.00
Scholarship	0.13	0.15	-0.02	0.73
Has prior debating experience	0.35	0.60	-0.24	0.00
Has won a debate prize before	0.32	0.11	0.21	0.00
Knows ChatGPT	0.47	0.32	0.15	0.07
Has used ChatGPT before	0.49	0.32	0.17	0.05
Baseline debate points (0-50)	33.74	25.33	8.40	0.00
Enjoyment (0-100) of speaking in public	71.80	64.52	7.29	0.13
Feels comfortable in debating language	0.66	0.50	0.16	0.05
Political position (1=left, 10=right)	6.32	6.65	-0.32	0.38
Polarised (0-100)	55.97	51.97	4.00	0.24
N	68	72		

Notes: The table shows summary statistics for the bottom and top 50% performers at baseline for the sample of students who registered to participate in the debating competitions. Column 1 reports the mean in the top 50% and Column 2 reports the mean in the bottom 50%. Column 3 reports the difference in the mean across the two groups, and Column 4 reports the p-value of a t-test of the equality in means across the two groups.

Table A3: Balance by scholarship status

	(1) Scholarship	(2) No scholarship	(3) Difference (1)-(2)	(4) p-value Col. (3)
Age	20.00	20.38	-0.38	0.30
Female	0.57	0.45	0.13	0.29
Father holds Master or higher degree	0.24	0.47	-0.23	0.05
Economics background	0.24	0.54	-0.30	0.01
Scholarship	1.00	0.00	1.00	.
Has prior debating experience	0.52	0.47	0.05	0.66
Has won a debate prize before	0.24	0.21	0.03	0.75
Knows ChatGPT	0.38	0.40	-0.02	0.84
Has used ChatGPT before	0.38	0.40	-0.02	0.89
Baseline debate points (0-50)	29.38	29.42	-0.03	0.98
Enjoyment (0-100) of speaking in public	79.79	65.95	13.84	0.04
Feels comfortable in debating language	0.95	0.52	0.43	0.00
Political position (1=left, 10=right)	4.57	6.83	-2.26	0.00
Polarised (0-100)	57.20	53.21	3.99	0.40
N	21	121		

Notes: The table shows summary statistics for the scholarship and non-scholarship recipients for the sample of students who registered to participate in the debating competitions. Column 1 reports the mean among those with a scholarship and Column 2 reports the mean among those with no scholarship. Column 3 reports the difference in the mean across the two groups, and Column 4 reports the p-value of a t-test of the equality in means across the two groups.

Table A4: Effect of GenAI on use of time and tools by student ability

	Time sufficient (0-100)		Useful tools (0-100)	
	(1)	(2)	(3)	(4)
	Low	High	Low	High
Treat	5.171	31.040*	35.839***	24.790
	(4.778)	(11.195)	(5.786)	(16.196)
Constant	52.877***	29.703**	47.201***	62.774***
	(5.106)	(9.193)	(6.468)	(10.362)
Mean dep. var.	64.25	38.75	46.60	56.55
SD dep. var.	26.78	29.24	33.47	34.84
R^2	0.11	0.40	0.32	0.13
Obs.	117	20	107	19

Significance levels are indicated by $^{+} < 0.1$, * $< .05$, ** $< .01$, *** $< .001$. The table shows results from regressions of a variable measuring self-reported valuations of whether time given to prepare debates was sufficient (0-100, with 0 being "not at all" and 100 being "plenty of time") and whether the tools given to prepare the debates were useful (0-100, with 0 being "not useful at all" and 100 "very useful") on a treatment dummy and debate day fixed effects, separately for low ability (non merit scholarship recipients) and high ability (merit scholarship recipients) students. The total number of individuals included in each regression differs between the columns because not all individuals answered all the questions at endline.

Table A5: Effect of GenAI on debate performance by own and rival's treatment status

	(1) Total points	(2) Winning debate
Treat	0.605	0.105
	(0.664)	(0.072)
Rival treated	0.026	-0.026
	(0.642)	(0.078)
Treat x Rival treated	0.140	-0.125
	(0.961)	(0.110)
Constant	13.908**	-0.851^{+}
	(4.943)	(0.510)
Mean dep. var.	28.67	0.48
SD dep. var.	4.53	0.50
R^2	0.21	0.11
Obs.	364	364

Notes: Significance levels are indicated by $^{+} < 0.1$, * < .05, ** < .01, *** < .001. The table shows results of regressing outcomes on treatment status and an interaction between own treatment status and that of ones rival. Both specifications include baseline debate points and the full set of controls as described in the notes to Table 2.

Table A6: Effect of GenAI on total points by debating position

	(1) Weak evidence	(2) Strong evidence
Treat	-0.190	1.471*
	(0.677)	(0.615)
Constant	16.321*	14.379*
	(6.761)	(5.733)
Mean dep. var.	28.12	29.37
SD dep. var.	4.53	4.45
R^2	0.29	0.26
Obs.	183	181

Significance levels are indicated by $^{+} < 0.1$, * < .05, ** < .01, *** < .001. The table shows results of regressing total debate points on treatment status, separately by topics that have a weak scientific evidence base and those that have a strong scientific evidence base. Both specifications include baseline debate points and the full set of controls as described in the notes to Table 2.

Table A7: Effect of GenAI on debate performance - judge-level regressions

	Debate points		Win	
	(1)	(2)	(3)	(4)
Treat	0.751	0.825*	0.009	0.021
	(0.473)	(0.408)	(0.035)	(0.032)
Constant	18.200***	6.746	0.041	-0.419
	(1.510)	(4.342)	(0.133)	(0.309)
Mean dep. var.	28.67	28.67	0.49	0.49
SD dep. var.	7.12	7.12	0.50	0.50
R^2	0.35	0.38	0.02	0.06
Obs.	1087	1087	1087	1087
Baseline score	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Notes: Significance levels are indicated by $^{+} < 0.1$, * $< .05$, ** $< .01$, *** $< .001$. This table shows results from regressions of equation 1, where the outcome variable is debate points (Columns 1 and 2) or a dummy equal to one if person i won in debate round r (Columns 3 and 4) evaluated by judge j. Specifications with controls (Columns 2 and 4) include the following control variables: age, gender, parental education, whether studies subject related to economics, prior debating experience, whether the students has prior experience using GenAI, whether the student is recipient of a scholarship, whether the student feels comfortable in the debating language, and ability measured by high school diploma grades and judge fixed effects. Standard errors are clustered at the individual level because each individual is observed between 6 and 9 times, depending on the number of debates they completed. The total number of individuals included in each regression is 141 out of 142 randomized individuals. One individual did not complete the baseline survey.

Table A8: Effect of GenAI on debating performance by scholarship status - judge-level regressions

	Debate points		Win	
	(1)	(2)	(3)	(4)
Treat	0.434	0.597	-0.002	0.033
	(0.496)	(0.440)	(0.037)	(0.035)
Scholarship	-0.744	-0.105	-0.057	0.016
	(1.014)	(0.905)	(0.083)	(0.070)
Treat x Scholarship	2.273^{+}	1.667	0.066	-0.087
	(1.195)	(1.360)	(0.094)	(0.096)
Constant	19.093***	8.213^{+}	0.078	-0.495
	(1.457)	(4.358)	(0.132)	(0.334)
Mean dep. var.	28.67	28.67	0.49	0.49
SD dep. var.	7.12	7.12	0.50	0.50
R^2	0.35	0.38	0.02	0.06
Obs.	1087	1087	1087	1087
Baseline score	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Significance levels are indicated by $^{+} < 0.1$, $^{*} < .05$, $^{**} < .01$, $^{***} < .001$. This table reports judge-level regressions of the effect of GenAI on debate performance by scholarship status. The outcome variable is debate points in Columns 1 and 2, and a dummy equal to one if person i won debate round r, according to judge j, in Columns 3 and 4. The coefficient on Treat captures the estimated effect of GenAI for students without a merit scholarship. The coefficient on Treat $\times$ Scholarship captures the additional effect of GenAI for merit-scholarship recipients relative to students without a scholarship. Specifications with controls (Columns 2 and 4) include the same controls as those reported in the notes to Table A7, including judge fixed effects. Standard errors are clustered at the individual level because each individual is observed between 6 and 9 times, depending on the number of debates completed and the number of judge evaluations. The total number of individuals included in each regression is 141 out of 142 randomized individuals; one individual did not complete the endline survey.

Table A9: Effect of GenAI on debating performance by above/below-median baseline performance - judge-level regressions

	Debate points		Win	
	(1)	(2)	(3)	(4)
Treat	-0.418	0.000	-0.007	0.035
	(0.610)	(0.608)	(0.051)	(0.053)
Top 50%	-1.126	-0.279	0.060	0.099
	(0.933)	(0.916)	(0.073)	(0.073)
Treat x Top 50%	2.310*	1.485	0.018	-0.042
	(0.948)	(1.019)	(0.074)	(0.077)
Constant	18.530***	9.094+	0.185	-0.278
	(2.028)	(4.609)	(0.177)	(0.318)
Mean dep. var.	28.67	28.67	0.49	0.49
SD dep. var.	7.12	7.12	0.50	0.50
R^2	0.35	0.38	0.02	0.06
Obs.	1087	1087	1087	1087
Baseline score	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

Significance levels are indicated by $^{+} < 0.1$, * $< .05$, ** $< .01$, *** $< .001$. This table reports judge-level regressions of the effect of GenAI on debate performance by baseline debating performance. The outcome variable is debate points in Columns 1 and 2, and a dummy equal to one if person i won debate round r, according to judge j, in Columns 3 and 4. The coefficient on Treat captures the estimated effect of GenAI for students below the median of the baseline debate-performance distribution. The coefficient on Treat $\times$ Top 50% captures the additional effect of GenAI for students above the median relative to students below the median. Specifications with controls (Columns 2 and 4) include the same controls as those reported in the notes to Table A7, including judge fixed effects. Standard errors are clustered at the individual level because each individual is observed between 6 and 9 times, depending on the number of debates completed and the number of judge evaluations. The total number of individuals included in each regression is 141 out of 142 randomized individuals; one individual did not complete the endline survey.

Table A10: Effects of GenAI support across debating subtasks

Debating Subtask	**Task Type**	**GenAI Role**	**Effect on Skill Gap**
Argument brainstorming	Inexpert implementation	AI can generate many plausible ideas on demand. Helps formulate examples.	Benefits lower-skill debaters, narrowing gaps in the quality and quantity of the potential pool of ideas available.
Structuring arguments	Inexpert implementation	AI provides well-organized outlines and clarity.	Disproportionately helps low-skill (improves their clarity), compressing differences in structure.
Opening speech	Mixed (requires some judgment in selection)	AI drafts content: human must choose the best arguments and present the arguments fluidly and persuasively.	High-skill use AI to enhance quality, low-skill may rely on AI verbatim. Moderate advantage remains with high-skill.
Refutation (rebuttal)	Expert judgment	AI not useful under time pressure; requires on-the-fly strategic thinking.	High-skill excel (AI can't assist much) – gap widens with GenAI support.
Rhetorical delivery & nuance	Expert judgment	AI can suggest phrasing, but effective delivery is human skill.	High-skill gain slightly (polish via AI), low-skill see little change, widening the gap in persuasion.

Table A11: Debating skill and relation to economic activities

Debate Skills	**Top Occupations According to O*NET**
Critical Thinking	Business Continuity Planners
	Quality Control Systems Managers
	Treasurers and Controllers
	Chief Sustainability Officers
	Financial Quantitative Analysts
Social Perceptiveness	Social and Community Service Managers
	Fundraising Managers
	Human Resources Managers
	General and Operations Managers
	Marketing Managers
Negotiation	Chief Executives
	Sales Managers
	Purchasing Managers
	Lodging Managers
	Wholesale and Retail Buyers
Persuasion	Chief Executives
	Marketing Managers
	Fundraisers
	Sales Engineers
	Chief Sustainability Officers

Table constructed from the ONET cross-functional skills dataset, showing the top five business-related occupations ranked by estimated skill level values reported by ONET.