

«GPT как соавтор» в командной работе студентов: анализ результатов внедрения в курс макроэкономики

Гипотеза комплементарности навыков AI и человека предполагает, что работа человека *совместно* с искусственным интеллектом обеспечивает более высокое качество решения практических задач, чем их навыки по отдельности (Steyvers et al., 2022; Rastogi et al., 2023; Hemmer et al., 2024). Включение AI как соавтора (AI as a collaborator) в групповую работу студентов способно улучшить навыки взаимодействия в команде, облегчить первичную генерацию идей при решении задач, стимулировать критический взгляд на приводимые аргументы и внутригрупповую дискуссию (Vizcaíno, 2005; Luckin et al., 2016; Lo, 2023; Atchley, 2024).

В этой разработке мы стремимся оценить, насколько существенно проявляется, и проявляется ли в принципе, комплементарность «человек-AI» в групповой работе студентов. Ключевой **мотивацией** для такой рефлексии служит необходимость понять, обладают ли студенты достаточным уровнем цифровой грамотности для того, чтобы эффективно интегрировать современные AI-инструменты (в первую очередь, GPT-модели) в работу команды.

Реализация: студентам второго курса Международной программы по экономике и финансам в рамках курса «Макроэкономика-1» еженедельно предлагается набор заданий для подготовки к семинарским занятиям в формате flipped classroom. Каждый набор заданий, помимо прочего, содержит два вопроса «верно/неверно», предполагающие развернутый и обоснованный ответ. Задания составлены таким образом, чтобы готовый ответ на них невозможно было найти в лекционных материалах или в учебнике, однако каждое из них в рамках изучаемой дисциплины предполагает однозначный ответ. В классе случайно выбранный студент из команды защищает решение устно, на основе чего вся команда получает оценку. Полный набор предложенных заданий представлен в Приложении 1.

В середине семестра студентам было предложено в дополнение к собственным ответам на вопросы «верно/неверно» использовать ChatGPT и GigaChat. Студентам предлагалось самостоятельно пройти путь от «наивного» промпта, дословно повторяющего задание, к «экспертному» взаимодействию с AI, включающему контекстуализацию запроса и использование материалов курса для пре-тренинга. По окончании работы с GPT-моделями студентам предлагалось выбрать наиболее содержательный ответ и загрузить его на интерактивную доску команды. Полный текст задания, предлагающего применить AI при подготовке решений, приведен в Приложении 2.

Качество ответов, сгенерированных AI, оценивалось преподавателями наравне с устными ответами студентов в ходе семинаров. Критерии оценивания включали:

- корректность и обоснованность утверждений
- специфичность и точность ответа на поставленный вопрос
- степень опоры на изученные в рамках курса экономические модели

Данные: для настоящего исследования использовались результаты работы в 8 академических группах, каждая из которых разбита на 2 или 3 команды по 4-7 человек. Для получения количественных выводов используются результаты проведения 8 семинарских занятий. Первая половина занятий не включала задания, требующие использование GPT для ответа на вопросы, а вторая – включала такой формат заданий. Оценки, полученные студентами за собственные и сгенерированные AI ответы, переведены в стобалльную шкалу. Для контроля общего уровня подготовки студентов в рамках каждой команды используется средняя итоговая оценка членов команды по курсу «математический анализ», полученная на первом курсе.

Гипотезы:

- 1) Использование GPT-моделей при решении задач повышает общий уровень подготовки команды к семинару благодаря комплементарности навыков студентов и AI
- 2) Применение AI-инструментов обеспечивает более равномерное усвоение материала внутри неоднородных групп студентов
- 3) Формат групповой работы способствует более быстрому освоению навыков AI-грамотности и повышает эффективность использования технологий искусственного интеллекта

Основные результаты:

Собранные данные демонстрируют, что качество ответов и аргументации не возросли после начала работы с AI при подготовке к семинару (см. Рисунок 1). Напротив, средний результат несколько снизился: за первые 4 семинара команды набрали в среднем 72,9 балла, а за последующие 4 – лишь 69,1.

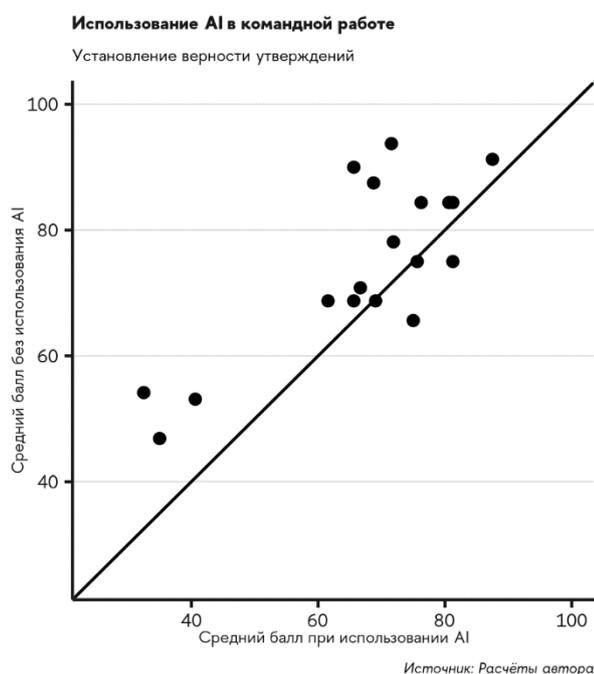


Рисунок 1. Средний балл команды за устный ответ: до и после использования AI

Одной из причин такой закономерности может служить постепенное усложнение материала курса по мере приближения к концу семестра. В результате положительный эффект от использования искусственного интеллекта при подготовке может компенсироваться более высокой сложностью вопросов, приводящей к ухудшению результатов во всех группах в целом.

Для формального тестирования гипотезы мы используем панельную регрессию со случайными эффектами, позволяющими контролировать влияние ненаблюдаемых факторов, ассоциирующихся со сложностью тех или иных заданий:

$$ScoreStudent_{i,j} = \beta_0 + \beta_1 \cdot PostAI_j + \beta_2 \cdot Calculus_i + \varepsilon_j, \varepsilon_j \sim N(0, \sigma_\varepsilon) \quad (1)$$

В уравнении (1) $ScoreStudent_{i,j}$ – оценка, полученная командой i при устной защите решения задачи j , $Calculus_i$ – средний балл членов группы по предмету «Математический анализ», который выступает как прокси общего уровня подготовки команды, а $PostAI_j$ – дамми-переменная, принимающая значение 1 для семинаров, при подготовке в которых студентам

предлагалось использовать GPT. Знак и величина коэффициента β_1 могут свидетельствовать о характере связи между использованием искусственного интеллекта студентами и результатом их работы.

Оценка коэффициентов модели (см. Таблицу 1) не позволяют выявить статистически значимого влияния применения AI, а знак зависимости при этом отрицателен. **Таким образом, сам по себе факт применения AI при подготовке к семинару не оказывает существенного влияния на результат команды.**

Таблица 1. Результаты оценивания панельных регрессий

Зависимая переменная: Балл за ответ, данный студентами		
	Модель (1)	Модель (2)
Работа с AI (PostAI)	-3.477	-
Балл AI (ScoreAI)	-	-0.118***
Ср. оценка группы по матанализу (Calculus)	0.425**	0.443***
Константа	52.383***	51.603***
Случайные эффекты: задача	Да	Да
Случайные эффекты: команда	Нет	Да
R2	0.07	0.11
Число наблюдений	146	56

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

На втором этапе мы стремимся оценить, существует ли комплементарность между возможностями AI и навыками студентов в командной работе. Для этого мы сопоставляем баллы, полученные студентами за защиту собственных решений, и тем, какую оценку получил ответ на то же самое задание, сгенерированный GPT. Рисунок 2 показывает, что устные ответы студентов в большинстве случаев были оценены существенно выше, чем ответы AI.

Формальная оценка того, насколько полезным для общего результата команды было использование AI, проводится с помощью панельной регрессии со следующей спецификацией:

$$ScoreStudent_{i,j} = \beta_0 + \beta_1 \cdot ScoreAI_{i,j} + \beta_2 \cdot Calculus_i + \varepsilon_j + \vartheta_i, \quad \varepsilon_j \sim N(0, \sigma_\varepsilon) \quad \vartheta_j \sim N(0, \sigma_\vartheta), \quad (2)$$

В уравнении (2) $ScoreStudent_{i,j}$ – оценка, полученная командой i при устной защите решения задачи j , $Calculus_i$ – средний балл членов группы по предмету «Математический анализ», а $ScoreAI_{i,j}$ – оценка, полученная командой i за ответ на вопрос, полученный с помощью GPT. Случайные эффекты ε_j и ϑ_i дополнительно контролируют ненаблюдаемые характеристики задач и команд соответственно. Коэффициент β_1 позволяет судить о степени комплементарности между навыками студентов и возможностями AI. Если он положителен, то команды, инвестирующие время и усилия в работу с GPT для получения более чёткого ответа, получают «выгоду» от этого при устной защите собственных решений.

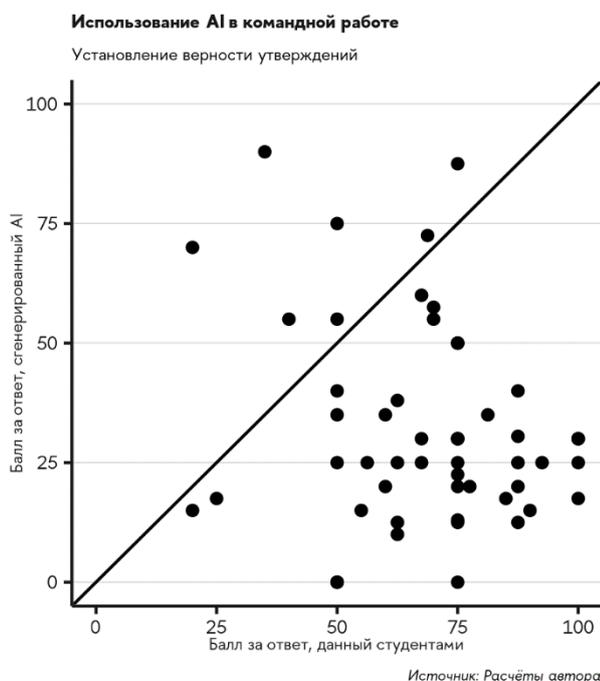


Рисунок 2. Качество ответов студентов и ответов, сгенерированных AI

Результаты оценки модели (2) приведены в Таблице 1: коэффициент β_1 статистически значим и отрицателен. **Это означает, что в ходе командной работы студенты не получают выгоды от комбинации сильных сторон AI и собственных навыков. Полученный результат свидетельствует об обратном: чем более качественное решение команда получает с помощью AI, тем хуже в среднем она показывает себя при защите собственного решения.**

Результаты количественного анализа в целом согласуются с впечатлением преподавателей и отзывами студентов об опыте работы с AI. Так, было замечено, что в большинстве групп работа с AI была делегирована наименее подготовленным студентам, в то время как более способные занимались решением задач самостоятельно. В результате студенты, взаимодействующие с GPT, работали автономно от остальной команды. Поскольку на этой дисциплине при устной защите решений преподаватель самостоятельно выбирает отвечающего, то участники команды, работавшие только с искусственным интеллектом, оказывались неспособны защитить итоговое командное решение.

Отсутствие видимых выгод от использования GPT для всей команды в тех (редких) случаях, когда за работу с AI отвечали сильные студенты, может объясняться тем, что им приходилось распределять время между взаимодействием с AI, подготовкой общекомандного решения и его объяснением остальным членам команды. В условиях ограниченных временных ресурсов последняя опция, вероятно, оказывалась наименее предпочтительной, что также ухудшало общий результат команды.

Анализ отдельных сгенерированных AI ответов показывает, что студенты не обладают достаточными навыками промптинга и зачастую не предоставляют GPT дополнительный контекст (например, материалы лекций или основные уравнения используемых моделей). Кроме того, не всегда делаются попытки скорректировать или улучшить ответы, в которых содержится неверная интерпретация применяющихся в курсе понятий. В качестве иллюстрации приведём пример задания, с которым работали студенты:

Consider a small open economy (SOE) with fixed prices and wages. Current account equals net exports. Capital and financial accounts are determined by the UIP condition. The balance of payments schedule is linear:

$$BP \equiv NX_0(Y^f) - z \cdot Y - d \cdot \frac{EP}{P^f} + cf \cdot \left(i - i^f + \frac{E^e - E}{E} - \rho \right) = 0$$

Goods market equilibrium requires: $Y = C_0 + c_1(1-t)Y + I_0 - b \cdot i + G_0 + NX_0(Y^f) - z \cdot Y - d \cdot \frac{EP}{P^f}$

The central bank targets money supply, so the money market clears when:

$$i = \frac{k}{h}Y + \frac{1}{h} \left(L_0 - \frac{M^S}{P} \right)$$

SECTION A: Artificial intelligence at work

Use the BEST prompting strategy and compare answers of the two LLMs (ChatGPT and GigaChat) to questions B1 and B2. Limit the answer by 4 sentences and ensure that there NO hallucinations, the proof is accurate and there is no ambiguity in wording: **[2 marks]**

Prompting strategy*	ChatGPT	GigaChat
Few-shot 'wise' prompt		

SECTION B: Basic definitions. True/False/Uncertain. Explain

HINT: Your verbal explanation should be concise and may include formulae and/or graphs if necessary. You can use the ideas from the AI answers and ensure that your answer adds value to the best response generated by AI. Highlight the difference and feel yourself as a genuine expert!

B1. [2 marks] Fiscal expansion causes higher investment in a SOE with imperfect capital mobility and flexible exchange rate.

B2. [2 marks] Balance budget fiscal contraction causes lower investment in a SOE with imperfect capital mobility and fixed exchange rate.

Рисунок 3. Пример задания «верно/неверно», включающего работу с AI

Ответы, данные AI при работе с утверждением B2 в двух командах, представлены на Рисунке 4.

Prompting strategy*	ChatGPT	GigaChat
Few-shot 'wise' prompt	<p>The impact of a balanced budget fiscal contraction on investment in a small open economy (SOE) with imperfect capital mobility and a fixed exchange rate is uncertain. Without specific context, it's unclear whether lower investment would result from a balanced budget fiscal contraction. Factors such as business expectations, the structure of the SOE, the sector, and overall economic conditions play a crucial role in determining the actual outcome.</p>	<p>The statement "Fiscal contraction to achieve budget balance leads to lower investment in a public company with limited capital mobility and a fixed exchange rate" may be true or false depending on the specific circumstances. Fiscal contraction may reduce overall demand in the economy, which in turn may reduce investment. Limited capital mobility and a fixed exchange rate can also affect a firm's investment activity. However, without additional details and analysis of the specific situation, it is impossible to give a definitive answer.</p>

<p>Few-shot 'wise' prompt</p> <p>B2</p> <p>3</p>	<p style="text-align: center;">TRUE</p> <p>True. In a small open economy with imperfect capital mobility and a fixed exchange rate, a balanced budget fiscal contraction is likely to cause lower investment. This is because reducing government spending (fiscal contraction) would lead to a decrease in aggregate demand, potentially lowering output (Y). Consequently, this reduction in output could result in lower interest rates (i), which would discourage private investment (I), as depicted by the goods market equilibrium equation.</p>	<p style="text-align: center;">TRUE</p> <p>True. A balance budget fiscal contraction in a small open economy with imperfect capital mobility and fixed exchange rate would cause lower investment. This is because a reduction in government spending or an increase in taxes leads to a decrease in aggregate demand, which in turn reduces investment. Additionally, the fixed exchange rate prevents any changes in the exchange rate as a response to the fiscal contraction, further limiting the potential for investment to be stimulated through external factors.</p> <p style="text-align: right;">4</p>
--	---	--

Рисунок 4. Примеры решения задания, сгенерированного GPT-моделями

Поскольку особенности полученных ответов оказались типичными, остановимся на них подробнее. Во-первых, **наиболее частой проблемой полученных от AI ответов и причиной их низкой оценки было отсутствие конкретики**. На Рисунке 3 (обозначение «1») обе GPT оказываются неспособны оценить верность утверждения, ссылаясь на отсутствие конкретного контекста. При этом само задание содержало уравнения, описывающие поведение всех секторов экономики, которые могли бы использоваться для предоставления контекста для GPT. В ходе семинаров выяснилось, что многие команды не знали о возможности работы с уравнениями при взаимодействии с AI.

Другой отличительной чертой является **недостаток работы с неявными предпосылками и интерпретациями, которые содержатся в ответах искусственного интеллекта**. На Рисунке 3 в области «2» аббревиатура SOE подразумевает расшифровку “state-owned enterprises”, в то время как в задании предполагается работа с моделью “small open economy” (малая открытая экономика). Область «3» содержит ошибочный с точки зрения макроэкономической теории вывод: низкая ставка процента при прочих равных стимулирует инвестиции, а не сокращает их. В области «4» используется предпосылка о зависимости инвестиций от совокупного спроса в экономике, которая явно противоречит уравнению, приведенному в преамбуле задания. Такие ошибки свидетельствуют о том, что результаты работы с AI зачастую глубоко не осмыслились и не подвергались анализу и совместному обсуждению внутри группы.

Дизайн этого исследования и отсутствие контрольной группы не позволяет претендовать на полноценную идентификацию причинно-следственных связей. Полученные статистические оценки следует воспринимать скорее как пищу для размышления и дополнение к собственным наблюдениям. При этом результаты расчётов и анализа данных не противоречат качественным выводам, которые разделяют преподаватели, работавшие на курсе. Представляется, что способность студентов бакалавриата экономических программ самостоятельно работать с AI-инструментами и извлекать из них пользу не следует принимать как само собой разумеющееся. Напротив, наш опыт показывает, что попытка интегрировать работу с GPT в подготовку к групповым семинарским занятиям может ухудшать итоговый результат из-за усиления фрагментации и специализации внутри команды. Кроме того, некоторые студенты демонстрировали выраженный скептицизм по отношению к использованию возможностей искусственного интеллекта при подготовке. По нашему мнению, это отражало скорее недостаточную информированность и отсутствие релевантной практики работы с AI, чем предыдущий негативный опыт. Диффузия навыков работы с GPT внутри команды также не происходит сама по себе. Таким образом, для раскрытия потенциала, сокрытого в комплементарности между возможностями AI и человеческих навыков, необходимо целенаправленно формировать у студентов соответствующие компетенции – как в рамках курсов «Цифровой грамотности», так и при преподавании конкретных дисциплин.

Литература:

Atchley, P., Pannell, H., Wofford, K., Hopkins, M., Atchley, R. (2024). Human and AI collaboration in the higher education environment: opportunities and concerns. *Cognitive Research: Principles and Implications* 9(20)

Hemmer, P., Schemmer, M., Kuhl, N., Vossing, M., & Satzger, G. (2024). Complementarity in Human-AI Collaboration: Concept, Sources, and Evidence. *ArXiv*, abs/2404.00029.

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410–425.

Luckin, R., Holmes, W., Griffiths, M., Forcier, L. (2016). *Intelligence Unleashed: An argument for AI in Education*. Open Ideas; Pearson Education, London.

Rastogi, C., Leqi, L., Holstein, K., and Heidari, H. (2023). A taxonomy of human and ML strengths in decision-making to investigate human-ML complementarity. *Proceedings of the AAAI Conference on Human Computation and mCrowdsourcing*, 11(1):127–139.

Steyvers, M., Tejada, H., Kerrigan, G., and Smyth, P. (2022). Bayesian modeling of human–AI complementarity. *Proceedings of the National Academy of Sciences*, 119(11):1–7.

Vizcaíno, A. (2005). A Simulated Student Can Improve Collaborative Learning. *International Journal of Artificial Intelligence in Education*. 15(1), 3–40.

Приложение 1: задания типа «верно/неверно», предлагавшиеся студентам

Семинар 4 (Growth policy):

- B1. [1 mark] The 'zero-growth' proposal states that it is optimal to have unchanged output per worker to minimize environmental damage.
- B2. [1 mark] Economic growth cannot be sustained in a resource-rich country unless the share of exported raw materials is minimized to zero.

Семинар 5 (Redistribution):

- B1. [1 mark] In a planned economy with only lump-sum taxes higher marginal propensity to expend always leads to higher equilibrium output.
- B2. [1 mark] The AK model of economic growth predicts that a permanent change in the saving rate has transitory effect on an economy's growth rate.

Семинар 7 (IS schedule):

- B4. [1 mark] The IS schedule is a behavioural relation showing that firms need to increase their output when real interest rate falls.
- B5. [1 mark] Changes in inflation expectations leave the IS schedule intact.

Семинар 8 (IS-MP model):

- B1. [1 mark] The MP schedule is a behavioural relation showing how money demand parameters determine the effectiveness of monetary policy.
- B2. [1 mark] Changes in inflation leave the MP schedule intact.

Семинар 9 (AD-AS model):

- B1. [1 mark] The AD schedule is a behavioural relation showing that economic agents shrink their aggregate expenditure because prices rise.
- B2. [1 mark] The AS schedule is a behavioural relation showing firms' incentives to increase production when inflation rises since real interest rate falls and investments become more attractive.

Семинар 10 (NX):

- B1. [1 mark] The NX(Y) schedule is a behavioural relation showing how induced imports offset autonomous exports at each level of real exchange rate.
- B2. [1 mark] The BPIS schedule in (Y,i) space is a locus showing how goods and forex markets clear simultaneously for any given nominal exchange rate.

Семинар 11 (IS-LM-BP model):

- B1. [2 marks] Fiscal expansion causes higher investment in a SOE with imperfect capital mobility and flexible exchange rate.
- B2. [2 marks] Balance budget fiscal contraction causes lower investment in a SOE with imperfect capital mobility and fixed exchange rate.

Семинар 12 (IS-LM-BP model):

- B1. [2 marks] Permanent fiscal expansion has no real effects in a small open economy with perfect capital mobility and flexible exchange rate in the long run when prices and wages are fully flexible.

B2. [2 marks] Temporary fiscal contraction has no real effects in a small open economy with perfect capital mobility and flexible exchange rate in the short run when prices and wages are fixed.

Семинар 13 (Uemployment):

B1. [1 mark] Shortening the working week allows to share the same amount of total work between more workers, so fewer people will be unemployed. Provide at least three different arguments

B2. [1 mark] Monetary policy shocks can affect NRU eve if they are temporary

Приложение 2: формулировка задания, направленного на использование AI при подготовке к семинару

SECTION AI: Artificial intelligence at work.

Use the FOUR prompting strategies and compare the answers of LLMs (ChatGPT and GigaChat) to the questions B1 and B2: **[2 marks]**

<u>Prompting strategy*</u>	ChatGPT	GigaChat
Zero-shot 'naïve' prompt		
Zero-shot 'wise' prompt		
Few-shot 'naïve' prompt		
Few-shot 'wise' prompt		

- Zero-shot 'naïve' prompt is identical to the original question, 'as is' wording
- Zero-shot 'wise' prompt is modified to make the question more contextual
- Few-shot 'naïve' prompt uses extra materials (lecture notes, slides, additional information) as a short pretrain but the original question is left 'as is'
- Few-shot 'wise' prompt – combines extra material for pre-train and contextualizes question

SECTION B: Basic definitions. True/False/Uncertain. Explain

HINT: Your verbal explanation should be concise and may include formulae and/or graphs if necessary. You can use the ideas from the AI answers and ensure that your answer adds value to the best response generated by AI. Highlight the difference and feel yourself as a genuine expert!