

УДК 372.862

**МЕТОДИКА БЕЗОПАСНОГО
И ЭФФЕКТИВНОГО
ИСПОЛЬЗОВАНИЯ БОЛЬШИХ
ЯЗЫКОВЫХ МОДЕЛЕЙ ОБУЧЕНИЯ
ПРОГРАММИРОВАНИЮ
ШКОЛЬНИКОВ: РЕЗУЛЬТАТЫ
ПЕДАГОГИЧЕСКОГО
ЭКСПЕРИМЕНТА**

Аннотация. Представлены результаты апробации методики безопасного и эффективного использования больших языковых моделей (LLM) в обучении программированию школьников 7–8-х классов. Эксперимент проведен на базе центра дополнительного образования «ИТ-Куб Южный» (г. Южно-Сахалинск) с участием 16 учащихся (ЭГ и КГ по 8 человек, 6 занятий по 90 минут). Изменялись пять зависимых переменных. Статистическая обработка (U-критерий Манна — Уитни, t-критерий Стьюдента) показала значимый прирост ЭГ по контрольной работе ($p = 0,010$) и мотивации ($p = 0,001$). Показатели промпт-инжиниринга ЭГ выросли с 26 до 72% от максимума. По самостоятельности значимых различий не выявлено. Результаты частично подтверждают гипотезу и указывают на риск формирования зависимости от LLM.

Ключевые слова: большие языковые модели, LLM, обучение программированию, педагогический эксперимент, промпт-инжиниринг, самостоятельность, мотивация.



**Константин Андреевич
Миль,**
*педагог дополнительного
образования
ГАУ РЦОКОСО,
подразделение «ИТ-Куб
Южный»,
г. Южно-Сахалинск, Россия*

Как цитировать статью: Миль К. А. Методика безопасного и эффективного использования больших языковых моделей обучения программированию школьников: результаты педагогического эксперимента // Образ действия. 2026. Вып. 3 «Естественно-научное образование (лучшие практики)». С. 122–129.

Большие языковые модели (от OpenAI ChatGPT до Anthropic Claude и Sber GigaChat) за последние годы стали частью образовательной среды раньше, чем педагогическое сообщество успело выработать единое отношение к их использованию. В результате вопрос уже заключается не в том, будут ли школьники применять LLM в обучении, а в том, как сделать это использование безопасным и педагогически осмысленным. В обучении программированию возможности LLM выглядят особенно перспективно: модели способны генерировать код, объяснять алгоритмы, помогать в поиске ошибок и предлагать альтернативные решения. Это создает предпосылки для персонализации обучения и более быстрой обратной связи. Однако эффективность таких инструментов напрямую зависит от методики их применения.

Как было показано в нашем предыдущем исследовании [1], неконтролируемое использование LLM связано с рядом рисков: снижением самостоятельности учащихся, формированием зависимости от готовых решений и возникновением «ложного ощущения понимания», при котором ученик уверен, что освоил материал, хотя основную интеллектуальную работу выполнила нейросеть. Парадоксально, что инструмент, предназначенный для поддержки мышления, при неудачном использовании может достаточно успешно это мышление заменять. Аналогичные выводы представлены в работе Стояновой и соавт. [5]: свободное использование ChatGPT без педагогической структуры не приводит к значимому росту академической успеваемости. Авторы подчеркивают, что положительный эффект возникает только при сочетании LLM с методиками, развивающими критическое мышление, а не механическое копирование решений. В то же время систематический обзор и метаанализ 69 исследований [3] зафиксировал положительное влияние ChatGPT на академические результаты, мотивацию и развитие мышления высшего порядка. Однако большинство работ выполнено в контексте высшего образования. Это важно, поскольку результаты исследований со студентами нельзя напрямую переносить на подростков с учетом возрастных особенностей обучения.

Цель исследования — экспериментальная проверка влияния разработанной методики безопасного и эффективного использования LLM на образовательные результаты учащихся 7–8-х классов при обучении программированию на языке Python.

Гипотеза исследования состоит в том, что систематическое применение методики, основанной на предварительном самостоятельном решении задач с последующим критическим анализом ответов LLM, позволит повысить уровень освоения учебного материала, навыки анализа кода, мотивацию и качество взаимодействия с языковыми моделями по сравнению с традиционным обучением без использования LLM.

Описание методики

Разработанная методика основана на принципе отсроченного использования LLM и предполагает, что нейросеть подключается не в начале решения задачи, а только после попытки самостоятельного выполнения. Такой подход позволяет сохранить за учеником основную интеллектуальную работу, а LLM использовать как инструмент анализа, а не как бесконечный генератор готовых ответов.

Работа с учебной задачей строится по 6-шаговому алгоритму: анализ условия без использования LLM, самостоятельное решение задачи, обращение к нейросети с конкретным запросом, сравнение собственного решения с ответом LLM, доработка кода и рефлексия. Принципиальное отличие методики от свободного использования LLM заключается в обязательном этапе самостоятельного решения. Нейросеть в данном случае выступает не заменой мышления, а объектом для критического анализа, так как ученик сначала решает задачу сам, а уже затем проверяет, действительно ли искусственный интеллект знает, о чем говорит.

Данный подход частично перекликается с системой НуроCompass [4], где LLM используется в роли «обучаемого агента»: студент помогает виртуальному персонажу находить ошибки в коде, формируя критическое отношение к ответам нейросети. В нашей методике эта идея расширяется: учащийся не только анализирует ошибки, но и сопоставляет различные варианты решения, оценивает их качество и формулирует собственные выводы.

Методика реализуется при постоянном педагогическом сопровождении. Преподаватель наблюдает за процессом работы, фиксирует качество запросов и уровень самостоятельности учащихся, при необходимости корректируя взаимодействие с LLM. Такой подход соответствует позиции, сформулированной в исследовании [6]: образовательная ценность языковых моделей определяется не самой технологией, а методикой ее применения и способностью поддерживать метакогнитивные навыки учащихся.

Организация эксперимента

Педагогический эксперимент проводился в апреле — мае 2026 года на базе «ИТ-Куб Южный» (г. Южно-Сахалинск). В исследовании приняли участие учащиеся 7–8-х классов в возрасте 14–15 лет, изучающие основы программирования на Python (циклы, списки, функции).

Участники были разделены на две группы по восемь человек. Экспериментальная группа (ЭГ) обучалась с использованием разработанной методики и LLM, контрольная группа (КГ) — по традиционной методике без использования языковых моделей. Тематика занятий и объем изучаемого материала в обеих группах были одинаковыми. Всего было проведено шесть занятий продолжительностью по 90 минут. В рамках исследования оценивались пять зависимых переменных:

- 1) уровень освоения учебного материала (контрольная работа, максимум — 21 балл);
- 2) навыки анализа кода (задания на объяснение, поиск ошибок и сравнение решений);
- 3) самостоятельность учащихся (карта наблюдения, шкала 0–3);
- 4) навыки промпт-инжиниринга (карта наблюдения, шкала 0–4, только для ЭГ);
- 5) учебная мотивация (анкета Ликерта из 12 утверждений, максимум — 60 баллов).

Для статистической обработки результатов использовались U-критерий Манна — Уитни и t-критерий Стьюдента. Уровень статистической значимости принимался равным $p < 0,05$.

Контрольная работа

На входном тестировании группы показали сопоставимые результаты: средний балл ЭГ составил 9,1, КГ — 8,8 из 21 возможного, что позволяет считать группы исходно однородными. По итогам эксперимента средний результат ЭГ увеличился до 13,6 балла, тогда как в КГ — до 10,8. Прирост в экспериментальной группе составил +4,5 балла против +2,0 в контрольной. Статистический анализ показал значимость различий ($U = 56,5$, $p = 0,010$; $t = 3,211$, $p = 0,006$). Наиболее выраженный рост результатов в ЭГ наблюдался в заданиях, связанных с анализом кода: объяснением логики работы программы, поиском ошибок и сравнением различных решений. Это особенно показательно, поскольку именно эти навыки обычно считаются наиболее уязвимыми при бесконтрольном использовании LLM.

Мотивация

На начальном этапе уровень мотивации в группах также был сопоставим: 38,1 балла в ЭГ и 37,0 — в КГ из 60 возможных. После завершения эксперимента показатель ЭГ вырос до 44,2 балла, тогда как в КГ — до 38,9. Прирост составил +6,1 балла в экспериментальной группе и +1,9 — в контрольной. Различия оказались статистически значимыми ($U = 63,0$, $p = 0,001$; $t = 4,338$, $p < 0,001$). Наибольшие изменения были зафиксированы в блоке, связанном с отношением учащихся к цифровым инструментам и интересом к процессу обучения.

Промпт-инжиниринг

Средний показатель качества запросов к LLM в экспериментальной группе вырос с 3,1 из 12 возможных баллов на втором занятии до 8,6 на пятом. В процентном соотношении рост составил с 26 до 72% от максимального значения. По мере прохождения занятий учащиеся постепенно переходили от общих и малоэффективных запросов к более точным формулировкам,

ориентированным на анализ конкретной проблемы. Наблюдалось снижение числа запросов, направленных на получение готового решения, и увеличение доли запросов, связанных с объяснением ошибок и логики работы программы. В одном из случаев ученик самостоятельно обнаружил логическую ошибку в ответе LLM, что можно рассматривать как показатель формирования более критического отношения к результатам нейросети.

Самостоятельность

Показатели самостоятельности были пересчитаны в проценты от максимального значения для удобства сопоставления. На занятиях 2–3 экспериментальная группа демонстрировала более высокие результаты (66,7 и 55,2%) по сравнению с контрольной группой (55,6 и 47,5%). Однако на занятиях 4–5 различия практически исчезли: показатели ЭГ составили 68,1 и 70,8%, КГ — 69,4 и 72,2%.

Полученные результаты по контрольной работе и мотивации в целом согласуются с данными метаанализа [3], где было зафиксировано положительное влияние ChatGPT на академические результаты и учебную мотивацию. При этом авторы отдельно отмечают возможный эффект новизны: часть роста мотивации может объясняться не только самой методикой, но и тем, что школьники работают с технологией, вокруг которой последние несколько лет строится примерно половина цифрового информационного шума человечества. В таких условиях сохранять к ней нейтральное отношение довольно сложно.

Рост навыков промпт-инжиниринга в экспериментальной группе с 26 до 72% соотносится с результатами исследования Клерка и соавт. [2], посвященного ИИ-грамотности школьников. Авторы показали, что даже краткосрочное обучение взаимодействию с LLM приводит к заметным изменениям в поведении учащихся: они начинают чаще задавать уточняющие вопросы, критичнее относиться к ответам модели и реже воспринимать их как автоматически истинные. Полученные нами результаты демонстрируют схожую динамику: по мере прохождения занятий запросы учащихся становились менее абстрактными и все больше напоминали попытку действительно разобраться в проблеме, а не просто получить готовый код максимально коротким путем, желательно еще до того, как придется читать условие задачи.

Наиболее интересными представляются результаты по самостоятельности. На первых этапах эксперимента учащиеся ЭГ демонстрировали более высокую вовлеченность, однако в дальнейшем различия между группами практически исчезли. Подобная динамика частично совпадает с выводами в других статьях [5], где отмечается, что при длительном использовании ChatGPT учащиеся начинают постепенно перекладывать часть мыслительной работы на модель.

В этом контексте особенно показательны результаты исследования Валле Торре и соавт. [6]. Авторы обнаружили отрицательную корреляцию между количеством запросов, ориентированных на получение готового решения, и итоговыми результатами обучения. При этом простое сокращение подобных запросов не приводило автоматически к росту успеваемости. Полученные нами данные позволяют предположить, что ключевую роль играет не сам запрет на использование LLM, а структура взаимодействия с нейросетью. Если ученик сначала самостоятельно пытается решить задачу, а уже затем использует модель для анализа и сравнения решений, LLM начинает работать как инструмент рефлексии.

Схожий принцип реализован и в системе НуроCompass [4], где LLM выступает не источником решений, а объектом анализа: студенту необходимо искать ошибки в ответах модели. Несмотря на ограниченный масштаб исследования, результаты подтверждают общий вывод, который прослеживается и в нашей работе: языковые модели оказываются наиболее полезными тогда, когда ученик сохраняет активную позицию в процессе обучения.

Таким образом, результаты эксперимента показывают, что педагогическая эффективность LLM определяется не самим фактом их использования, а способом интеграции в учебный процесс. При отсутствии структуры нейросеть действительно может сводить обучение к механическому копированию решений. Однако при наличии педагогического контроля, этапов самостоятельного решения и рефлексии LLM способны становиться инструментом развития аналитического мышления. По крайней мере, до того момента, пока очередная модель снова не пообещает «полностью изменить образование», как это уже происходило примерно с каждым новым цифровым инструментом за последние 20 лет.

Ограничения

Исследование имеет ряд ограничений, которые необходимо учитывать при интерпретации результатов. Прежде всего речь идет о малом объеме выборки ($n = 8$ в каждой группе), что ограничивает возможность широкого обобщения полученных данных. Впрочем, для исследований в области применения LLM в образовании подобные масштабы пока скорее правило, чем исключение [4; 6]: технологии развиваются значительно быстрее, чем успевают накапливаться большие и устойчивые выборки.

Дополнительным ограничением является отсутствие отсроченного контроля. Исследование фиксирует изменения в рамках относительно короткого периода, однако не позволяет оценить устойчивость полученного эффекта в долгосрочной перспективе. Это особенно важно в контексте мотивации, поскольку, как отмечается в метаанализе [4], часть положительного эффекта ChatGPT может объясняться фактором новизны. Иными словами, сложно заранее определить, сохраняется ли высокий уровень

вовлеченности после того, как взаимодействие с нейросетью перестает восприниматься как технологическое событие.

Нельзя исключать и влияние эффекта Хоторна: учащиеся могли демонстрировать более высокую активность просто из-за осознания участия в эксперименте. Для школьной среды это вполне типично: иногда сам факт того, что «на уроке разрешили нейросеть», уже становится отдельным мотивационным фактором.

Следует учитывать и особенности используемых методов измерения. Мотивация оценивалась с помощью самоотчетных анкет, тогда как исследование Клерка и соавт. [2] показывает, что субъективная уверенность учащихся не всегда коррелирует с реальными образовательными результатами. Кроме того, наблюдение за самостоятельностью учащихся и качеством промпт-инжиниринга проводилось одним исследователем, что потенциально увеличивает риск субъективности при оценке результатов.

Результаты апробации частично подтверждают гипотезу. По четырем из пяти переменных зафиксирована положительная динамика ЭГ: значимый прирост баллов контрольной работы ($p = 0,010$), мотивации ($p = 0,001$), рост навыков анализа кода и промпт-инжиниринга (с 26 до 72%). По самостоятельности различия не выявлены, что согласуется с данными [5; 6] о риске формирования зависимости от LLM.

Сравнение с зарубежными исследованиями позволяет заключить, что структурированная методика (6-шаговый алгоритм) обеспечивает более выраженный положительный эффект, чем свободное использование LLM [5] или изолированные интервенции по регуляции запросов [6]. Ключевым фактором является сочетание самостоятельного решения, критического анализа и рефлексии.

Практическая значимость: методика применима в любом образовательном учреждении без специального оборудования — достаточно доступа к бесплатной LLM. Полноценный эксперимент с увеличенной выборкой и отсроченным контролем планируется на 2-м курсе магистратуры.

Список литературы

1. Миль К. А. Потенциал и риски использования больших языковых моделей в школьном обучении программированию. 2026.
2. Clerc O., Abdelghani R., Desvaux C., et al. Teaching Students to Question the Machine: An AI Literacy Intervention Improves Students' Regulation of LLM Use in a Science Task // arXiv. 2026. doi: 10.48550/arXiv.2604.01955.
3. Deng R., Jiang M., Yu X., et al. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies // Computers & Education. 2025. P. 227.
4. Ma Q., Shen H., Koedinger K. R., et al. HypoCompass: Large-Language-Model-based Tutor for Hypothesis Construction in Debugging for Novices // arXiv. 2023. doi: 10.48550/arXiv.2310.05292.

5. Stoyanova D., Stoyanova-Petrova S.; Shotarova S., et al. ChatGPT in Programming Education: An Empirical Study on Its Impact on Student Performance, Creativity, and Teamwork // Education Sciences. 2026. Vol. 16, no. 1. P. 19.
6. Valle Torre M., Specht M., Oertel C. LLM Chatbots in High School Programming: Exploring Behaviors and Interventions // arXiv. 2025. doi: 10.48550/arXiv.2511.18985.