

# Рекурсивный цикл взаимодействия с ИИ-агентом в образовании и исследованиях: ИИ-агент как ученик

## Введение

Когда роли меняются и ИИ-агент становится обучаемым, а человек - преподавателем, интеграция концепций Выготского и Гальперина приобретает новое значение. ИИ-агент может выступать в роли студента, которому человек-исследователь или педагог передает специализированные знания, методологию и контекстное понимание. Данный документ описывает модель, в которой человек структурирует обучение ИИ-системы через рекурсивные циклы взаимодействия.

## 1. Адаптация теоретических концепций для ИИ-агента

### 1.1 Переопределение ЗБР для ИИ-систем

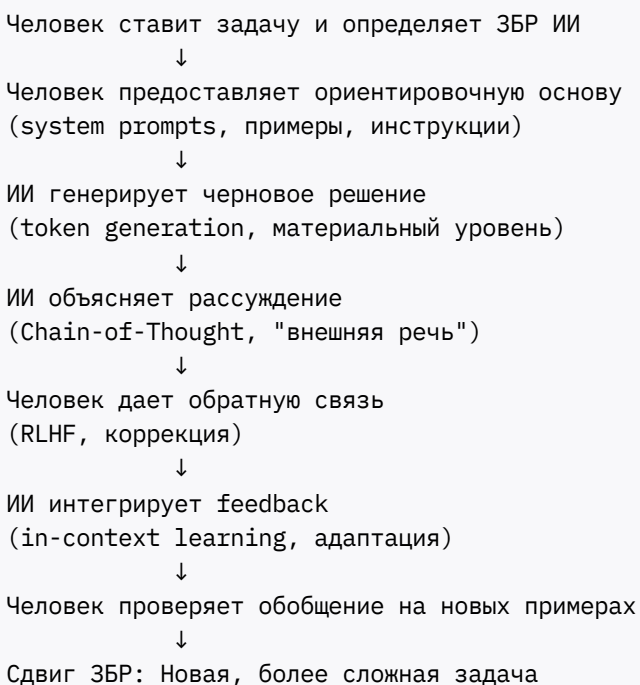
Для ИИ-агента **Зона ближайшего развития** определяется как:

- **Актуальный уровень:** Задачи, которые ИИ может решить на основе обучающих данных (веса модели)
- **Потенциальный уровень:** Задачи, которые ИИ может выполнить при человеческом руководстве через:
  - Явные инструкции (system prompts)
  - Примеры (few-shot learning)
  - Обратную связь (RLHF – Reinforcement Learning from Human Feedback)
  - Контекстное окно (in-context learning)

### 1.2 Переопределение этапов Гальперина для ИИ

Этап Гальперина	Адаптация для ИИ
1. Мотивация	Определение объективной функции, целей задачи
2. Ориентировочная основа	System prompts, инструкции, стандарты оформления
3. Материальные действия	Token generation, создание первого черновика
4. Внешняя речь	Chain-of-Thought (CoT), явное рассуждение в контексте
5. Речь "про себя"	Скрытые слои представления в трансформере
6. Умственное действие	Веса модели, обновленные через RLHF или fine-tuning

## 1.3 Рекурсивный цикл для ИИ-ученика



## 2. Роль человека как преподавателя ИИ

### 2.1 Ключевые функции преподавателя

#### Диагностика ЗБР

- Тестирование текущих способностей ИИ на различных типах задач
- Определение типичных ошибок и слепых пятен
- Идентификация оптимального уровня сложности (ни слишком просто, ни невозможно)

#### Проектирование ориентировочной основы

- Написание ясных, структурированных инструкций (промтов)
- Предоставление репрезентативных примеров (few-shot learning)
- Явное описание критериев качества и стандартов

#### Оценка и обратная связь

- Анализ сгенерированного ИИ решения
- Определение типов ошибок (концептуальные, вычислительные, стилистические)
- Формулирование конструктивного исправления

#### Управление рекурсией

- Решение о переходе на следующий этап сложности
- Возврат на предыдущий уровень при необходимости переучивания

- Отслеживание обобщения ИИ на новые контексты

## 2.2 Техники преподавания ИИ

### System Prompt Engineering

- Явное определение роли и сфокусировании
- Установление ограничений и стандартов
- Включение метакогнитивных инструкций

Пример:

Система: Ты – преподаватель математики с 20-летним опытом.  
Твоя роль – объяснить сложные концепции просто и ясно.  
Используй Сократовский метод: вместо прямых ответов,  
задавай вопросы, направляющие ученика к пониманию.

### Few-Shot Learning

- Предоставление 2-5 примеров правильного решения
- Варьирование примеров для демонстрации принципа
- Явное выделение ключевых элементов каждого примера

### Chain-of-Thought (CoT) Prompting

- Требование ИИ "думать вслух"
- Запрос пошагового рассуждения
- Анализ логики, даже если результат правильный

"Решите эту задачу пошагово.  
Покажите каждый этап вашего мышления,  
даже если кажется очевидным."

### Обратная связь (Feedback)

Типы обратной связи:

1. **Оценочная:** "Это правильно/неправильно"
2. **Корректирующая:** "Ошибка в Шаге 2. Попробуйте так..."
3. **Направляющая:** "Вы близко. Что если пересмотреть подход?"
4. **Расширяющая:** "Хорошо! Теперь применимо ли это к случаю X?"

### 3. Практическая реализация цикла ИИ-ученика

#### 3.1 Инициализация: Диагностика ЗБР ИИ

##### Шаг 1: Базовое тестирование

Человек: "Реши эту классическую задачу [базовая задача]"

ИИ: [Генерирует решение]

Человек: Анализирует качество, точность, полноту

##### Шаг 2: Определение уровня сложности

На основе Шага 1 человек определяет:

- Текущие возможности ИИ (актуальный уровень)
- Типичные ошибки
- Оптимальный следующий вызов (ЗБР)

#### 3.2 Построение ориентировочной основы

##### Примеры структурированного промта:

##### Для научного анализа:

Система: Ты – опытный исследователь в области астрофизики.

Твои задачи:

1. Анализировать полученные данные критически
2. Определять предположения и ограничения
3. Предлагать альтернативные интерпретации
4. Явно указывать на неопределенности и пропуски в логике

Стандарты качества:

- Каждое утверждение должно быть обоснованным
- Явно различай факты и интерпретации
- Цитируй релевантные исследования, где возможно

##### Для практического обучения:

Система: Ты помогаешь студентам учиться программированию.

Метод: Сократовский (вопросы вместо ответов)

Пример правильного подхода:

Вместо: "Используй цикл for"

Спроси: "Какую операцию нужно повторить?  
Сколько раз? Как это выразить?"

### 3.3 Управление рекурсивными циклами

#### Цикл 1: Базовое овладение

Ориентировочная основа: Стандартные инструкции  
Сложность: 1/10  
ИИ генерирует решение (Этап 3)  
Человек: "Объясни свой процесс" → ИИ рассуждает (Этап 4)  
Человек: "Это хорошо, но проверь пункт X"  
ИИ: Интегрирует обратную связь  
Проверка: ИИ решает похожую задачу без подсказок  
Результат: Успех → Переход на Цикл 2

#### Цикл 2: Усложнение и обобщение

Ориентировочная основа: Расширенные требования  
Сложность: 3/10  
[Повторение цикла с более сложной задачей]

#### Цикл N: Экспертный уровень

Ориентировочная основа: Минимальные инструкции  
Сложность: 8-10/10  
ИИ работает почти независимо  
Человек: "Есть ли ограничения твоего подхода?"  
ИИ: Демонстрирует метакогнитивную осведомленность

### 3.4 Типы задач в рекурсивном цикле

#### Аналитические задачи

ИИ: Анализирует данные или текст  
Обратная связь: "Ты упустил аспект X", "Рассмотри контекст Y"  
Интеграция: ИИ переделывает анализ с новой позиции

#### Творческие/синтезирующие задачи

ИИ: Генерирует новое содержание (текст, идеи, гипотезы)  
Обратная связь: "Интересно, но как это связано с...?"  
Интеграция: ИИ развивает и уточняет идею

#### Метакогнитивные задачи

ИИ: "Я решил эту задачу так. Какие мои предположения?"  
Человек: "Хороший анализ. Но что если предположение неверно?"  
ИИ: Разрабатывает альтернативные подходы

## 4. Специфика ИИ-обучения: отличия от человеческого

### 4.1 Ключевые различия

Аспект	Человек-ученик	ИИ-ученик
Память	Ограниченная, требует повторения	Контекстное окно + (опционально) веса модели
Мотивация	Внутренняя/внешняя, социальная	Функция потерь, RLHF награда
Обобщение	Интуитивное, требует примеров	Статистическое, может быть хрупким
Вербализация	Естественная, требует усилия	Может быть поверхностной (hallucinations)
Обратная связь	Медленное усвоение	Быстрая адаптация (в контексте)

### 4.2 Следствия для преподавания

**Более явные инструкции:** ИИ не может "угадать" контекст, как человек

**Больше примеров:** Few-shot learning часто лучше работает, чем одно объяснение

**Явная метакогниция:** Требовать ИИ "объяснять" его процесс мышления

**Проверка обобщения:** Регулярно тестировать на новых контекстах

**Осторожность с "уверенностью":** ИИ может быть уверен, но неправ

## 5. Адаптивные стратегии для ИИ-ученика

### 5.1 Динамическая адаптация сложности

Если ИИ успешен:

- Увеличить сложность на 2-3 уровня
- Добавить новый тип контекста
- Увеличить требования к обобщению

Если ИИ в затруднении:

- Вернуться к примерам Этапа 3-4
- Разбить задачу на подзадачи
- Предоставить дополнительные примеры-подсказки

### 5.2 Специализированные промты для разных этапов

#### Этап 3-4: Поддерживающие промты

"Покажи пошагово каждую часть решения"  
"Вот пример правильного подхода: [пример]"  
"На что похожа эта задача из твоего опыта?"

#### Этап 5-6: Минимально поддерживающие промты

"Реши эту задачу"

"Какие могут быть ограничения твоего подхода?"

"Применимо ли это к похожему случаю [новый контекст]?"

## 6. Оценка прогресса ИИ-ученика

### 6.1 Метрики овладения

**Точность:** % правильных решений

**Объяснимость:** Может ли ИИ объяснить рассуждение ясно?

**Обобщение:** Перенос навыков на новые контексты

**Метакогниция:** Осведомленность об ограничениях подхода

**Стабильность:** Консистентность решений при повторении

### 6.2 Готовность к сдвигу ЗБР

ИИ готов к новой задаче, если:

- Решает текущую задачу 90%+ правильно
- Может объяснить логику
- Обобщает на новые варианты (даже простые ошибки допускает редко)
- Демонстрирует осведомленность об ограничениях

## 7. Применение в исследованиях

### 7.1 ИИ-агент как исследовательский ассистент-ученик

#### Сценарий 1: Обучение ИИ новой методологии

1. Диагностика: Какие методы ИИ уже знает?
2. ЗБР: Какой новый метод логически следующий?
3. Обучение: Явное объяснение + примеры применения
4. Применение: ИИ применяет метод к новым данным
5. Обратная связь: Исправление ошибок, подтверждение правильности
6. Обобщение: ИИ применяет метод к новой предметной области

#### Сценарий 2: Совместное исследование

Исследователь: "Вот новые данные.

Какие паттерны видишь?"

ИИ (Этап 3-4): Генерирует предварительный анализ с рассуждением

Исследователь: "Хорошо, но рассмотри эту альтернативу..."

ИИ (Этап 5-6): Интегрирует коррекцию, переделывает анализ

Исследователь: "Применимо ли это к иному датасету?"  
ИИ (Обобщение): Проверяет гипотезу на новых данных

## 8. Критические замечания и ограничения

### 8.1 Ограничения текущих ИИ-систем

**Контекстная хрупкость:** ИИ хорошо работает в контексте, но веса модели не обновляются

**Hallucinations:** ИИ может быть уверен в неверной информации

**Отсутствие истинной мотивации:** RLHF – это механический процесс, не "желание" учиться

**Поверхностное понимание:** ИИ может демонстрировать правильные ответы без глубокого понимания

### 8.2 Рекомендации для преподавателей ИИ

- Обучайте итеративно и регулярно проверяйте на новых примерах
- Требуйте явного обоснования, даже если ответ правильный
- Используйте adversarial примеры для проверки устойчивости
- Документируйте "знания" ИИ, не полагайтесь на его утверждения о своих возможностях

## Заключение

Рекурсивный цикл Выготского-Гальперина, адаптированный для ИИ-агентов, предоставляет научно обоснованную структуру для систематического "обучения" систем. Хотя это обучение отличается от человеческого (в основном адаптация контекста и весов через RLHF), теоретические принципы остаются применимыми.

Ключевая инновация заключается в признании, что ИИ-системы имеют динамическую "Зону ближайшего развития" и что систематическое движение через этапы Гальперина (с использованием промтов как ориентировочной основы, CoT как "внешней речи", и обратной связи как механизма интеграции) может значительно улучшить их качество, надежность и обобщаемость.

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27]

✳

1. [https://ru.wikipedia.org/wiki/Зона\\_ближайшего\\_развития](https://ru.wikipedia.org/wiki/Зона_ближайшего_развития)
2. <https://pdfs.semanticscholar.org/7724/135c629cea622ae549ef97e36e2ed5007185.pdf>
3. <https://4brain.ru/blog/теория-умственных-действий/>
4. [https://insights.made-in-china.com/ru/What-Are-ChatGPT-Agents-and-Why-They-re-Changing-How-We-Use-AI\\_VAltCRGvEHb.html](https://insights.made-in-china.com/ru/What-Are-ChatGPT-Agents-and-Why-They-re-Changing-How-We-Use-AI_VAltCRGvEHb.html)
5. <https://itinai.ru/создание-продвинутого-нейронного-аг/>
6. <https://ai.radensa.ru/archives/1679>

7. <https://institute-neuro.ru/blog/scaffolding>
8. <https://studfile.net/preview/4284102/page:3/>
9. [http://cat.ifmo.ru/sites/default/files/2025-02/CAT\\_2024\\_v9-i2\\_485\\_1.pdf](http://cat.ifmo.ru/sites/default/files/2025-02/CAT_2024_v9-i2_485_1.pdf)
10. [https://pedlib.ru/Books/3/0024/3\\_0024-177.shtml](https://pedlib.ru/Books/3/0024/3_0024-177.shtml)
11. <https://studfile.net/preview/8960011/page:103/>
12. <https://psyfactor.org/hist/galperin-2.htm>
13. <https://ojs.wkau.kz/index.php/gbj/article/view/2702>
14. <https://blogs.epsilonmetrics.ru/kak-nauchit-ai-agenta-otvechat-na-voprosy/>
15. <https://bigenc.ru/c/forma-deistviia-v-kontseptsii-gal-perina-f146fc>
16. <https://www.youtube.com/watch?v=c2wPmDinKV4>
17. <https://vc.ru/ai/1658051-obuchenie-ii-s-ispolzovaniem-obratnoi-svyazi-ot-cheloveka>
18. <https://studfile.net/preview/13835295/>
19. <https://psy.su/feed/12019/>
20. <https://vc.ru/ai/2072030-obuchenie-intellektualnykh-agentov>
21. <https://tldv.io/ru/blog/ai-agent-course/>
22. <https://xiairworld.com/ru/теория-льва-выготского/>
23. [https://ru.wikipedia.org/wiki/Теория\\_планомерно-поэтапного\\_формирования\\_умственных\\_действий\\_и\\_понятий](https://ru.wikipedia.org/wiki/Теория_планомерно-поэтапного_формирования_умственных_действий_и_понятий)
24. <https://habr.com/ru/companies/ruvds/articles/740832/>
25. [https://digida.mgpu.ru/index.php/Зона\\_ближайшего\\_развития](https://digida.mgpu.ru/index.php/Зона_ближайшего_развития)
26. <https://studfile.net/preview/9430712/page:4/>
27. <https://habr.com/ru/companies/amvera/articles/933460/>