

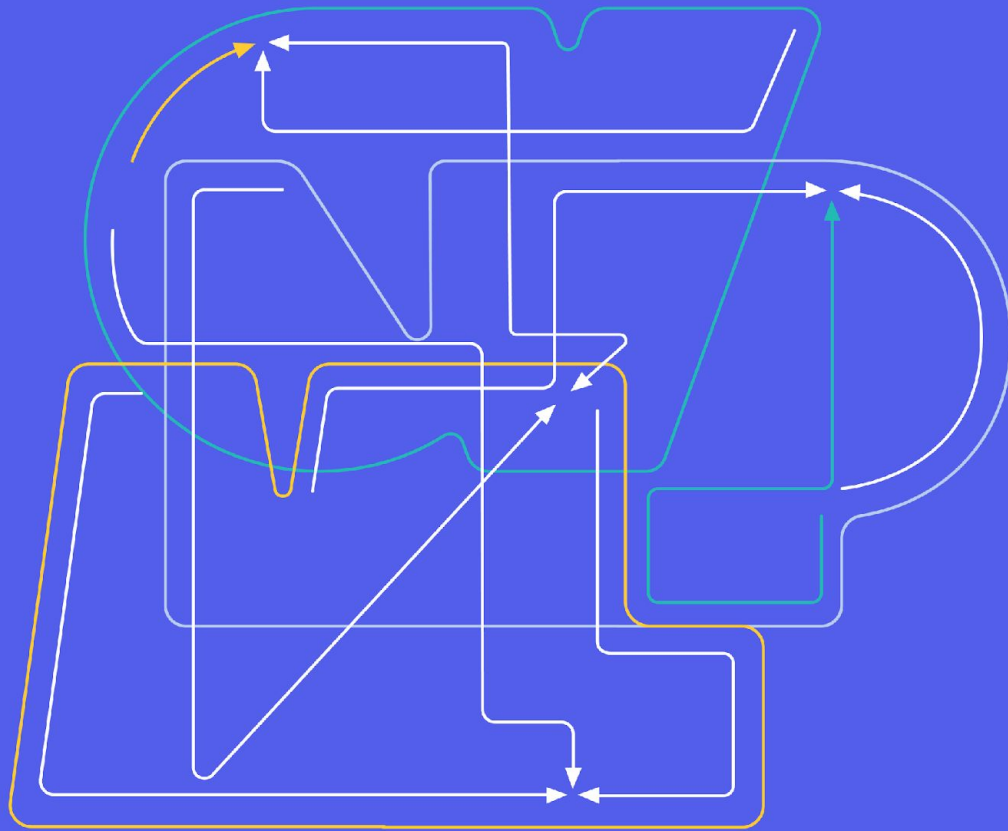


Национальный  
исследовательский  
Томский  
государственный  
университет

# Applied AlphaEvolve

Next-Gen LLMs

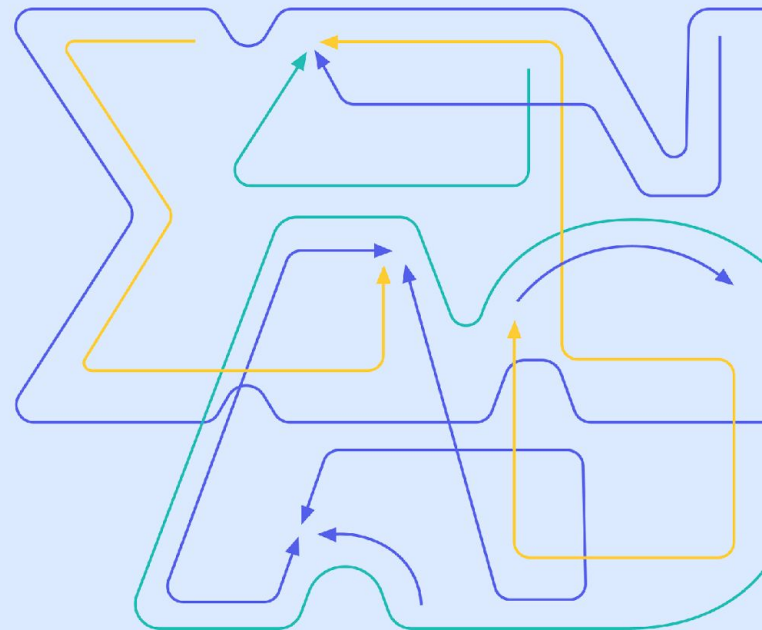
Проект Летней школы AIRI 2025, Томск



# 01

---

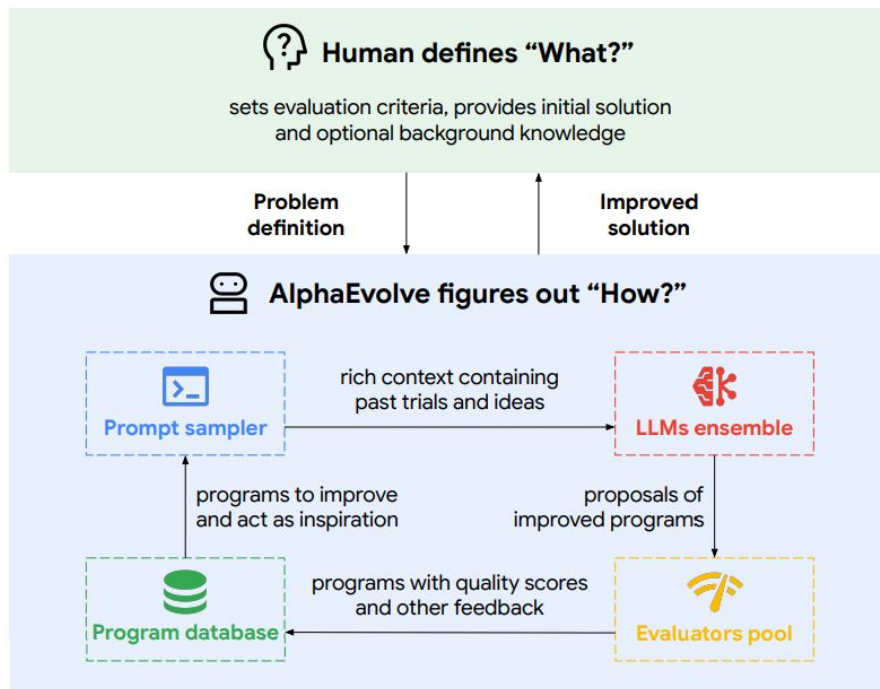
Что такое  
AlphaEvolve?



# AlphaEvolve

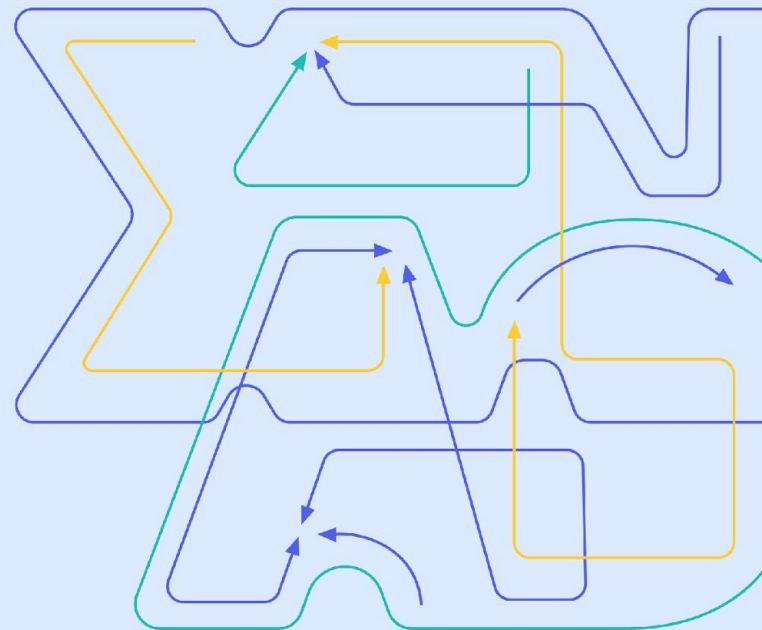
– это автономный агент, сочетающий возможности современных больших языковых моделей (LLM) с эволюционным поиском программ.

- ✓ Работает не по принципу «один запрос – один ответ», а как непрерывный цикл улучшений
- ✓ Превзошел алгоритм Штрассена
- ✓ Установил новый рекорд в задаче kissing number (упаковка сфер)
- ✓ Улучшил результат в задаче Эрдёша о перекрытии



# 02

## Задачи проекта

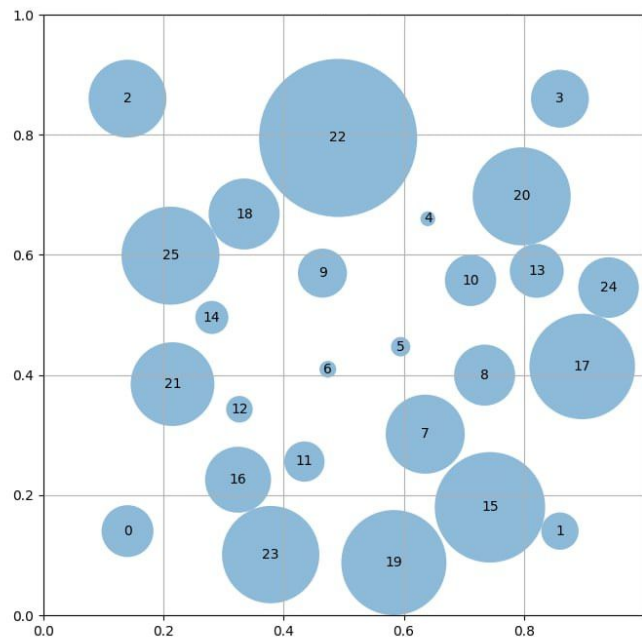


# Задачи

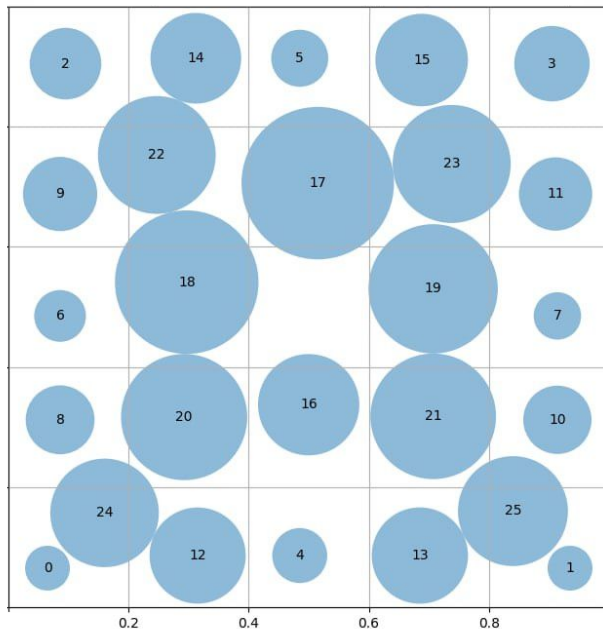
- Изучить архитектуру и внутреннюю логику существующих реализаций AlphaEvolve.
- Протестировать работу системы с различными LLM - от проприетарных больших до открытых маломощных, оценить её устойчивость и адаптируемость. Ответить на вопрос применимости такого подхода для open-source решений.
- Продумать возможные улучшения:
  - Концептуальные: улучшение стратегии мутаций, адаптация метрик к задаче
  - Визуальные: более наглядная визуализация процесса эволюции
  - Технические: расширение логирования, отслеживание траектории решений



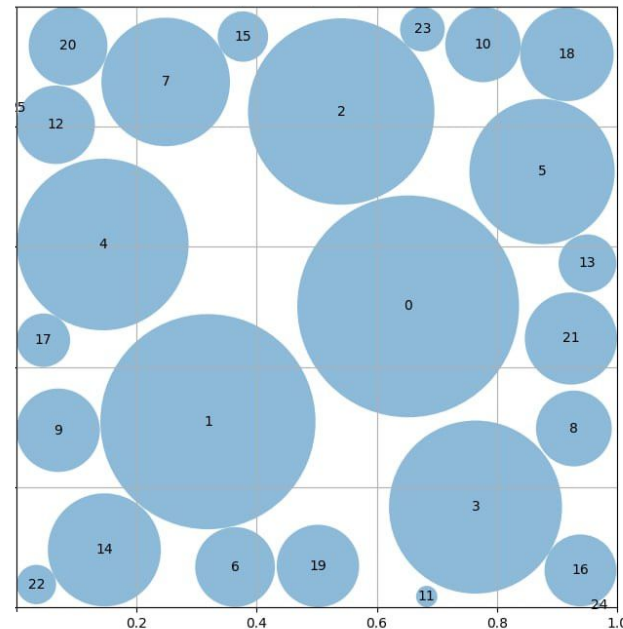
# Результаты проприетарных моделей в zero-shot режиме



(a) GPT-o3. Sum = 1.44



(b) Claude Sonnet 4 Thinking Sum = 1.95

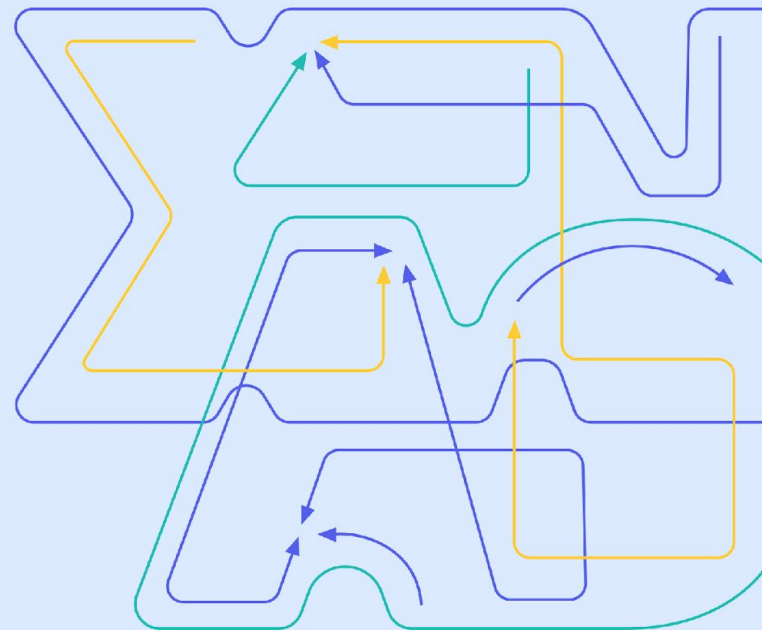


(c) GPT-o3 Pro. Sum = 2.03

Решается задача упаковки 26 кругов в единичный квадрат

# 03

## Обзор решений



# Сравнение реализаций AlphaEvolve

	OpenAlpha_Evolve	OpenEvolve	 MetaEvolve
Мутации	✓	✓	✓
Кроссинговер	—	✓	✓
Миграции	✓	✓	✓
Баланс случайности	✓	—	✓
Multi-Island	✓	✓	✓
MAP-Elites	—	✓	✓ ✓
Метрики	—	✓	✓ ✓
Асинхронность	✓	✓	✓
Удобность логирования	✓	✓	—
Масштабируемость	—	—	✓



# Сравнение реализаций AlphaEvolve

## Плюсы

## Минусы

### OpenAlpha\_Evolve

- + Поддержка Multi-Island
- + Модульность
- + Удобное логирование
- + Асинхронность

- Нет кроссинговера (только мутации и bug-fix)
- Слабая масштабируемость хранения
- Ограниченная метрика → ограниченная селекция

### OpenEvolve

- + LLM-ансамбль
- + Поддержка Multi-Island MAP-Elites
- + Чистая архитектура
- + Асинхронность
- + Логирование по каждому острову

- Нет случайных миграций, что ограничивает новые идеи
- Слабая масштабируемость хранения

### MetaEvolve



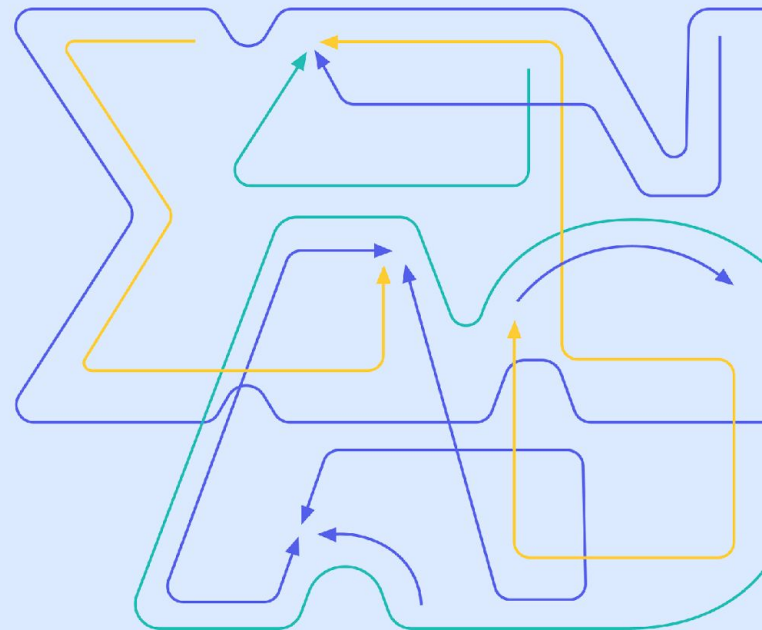
- + Поддержка Multi-Island MAP-Elites
- + Гибкая архитектура на основе DAG
- + Асинхронность на всех этапах
- + Поддержка Redis

- Много гиперпараметров
- Неудобное логирование

# 04

---

## Результаты



# Сравнение результатов исследуемых реализаций AlphaEvolve

Фреймворк	Модели	Вес модели	Метрика
<b>OpenAlpha_Evolve</b>	Qwen2.5 7B	1.0	1.811
<b>OpenEvolve</b>	mistral-small mistral-medium	0.8 0.2	2.321
	mistral-medium mistral-large	0.8 0.2	2.412
	Qwen3 14B (thinking) Qwen2.5 Coder 14B	0.5 0.5	2.618
	deepseek-v3-0324 gemini-2.0-flash-exp	0.2 0.8	2.620
<b>MetaEvolve</b>	Qwen 2.5 Coder 32B	0.2	1.883
	devstral-small	0.2	
	gemini-2.0-flash-exp	0.2	
	qwen3-235b-a22b (thinking)	0.2	
	deepseek-r1 (thinking)	0.1	
	llama-3.3-70b	0.1	

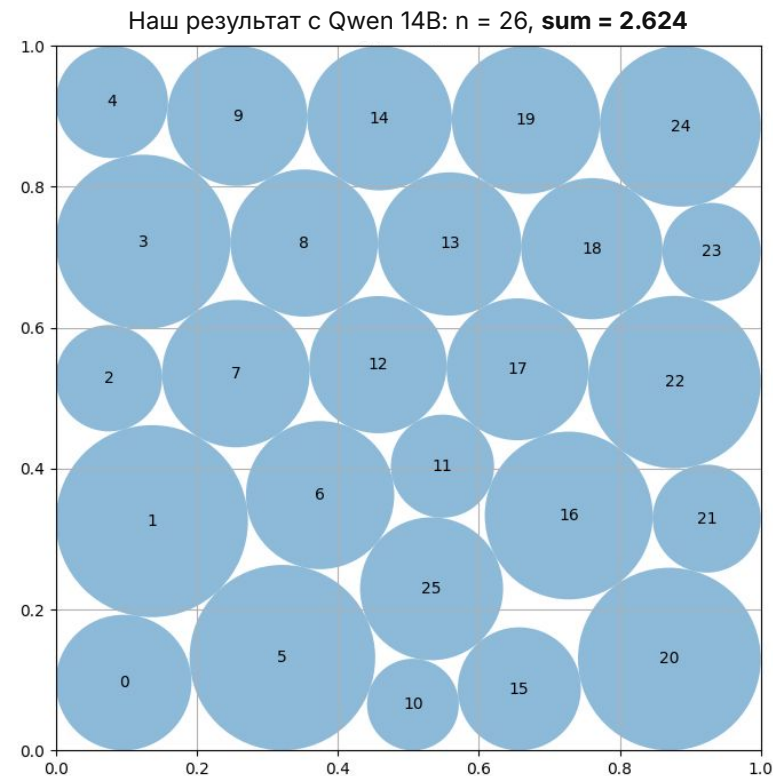
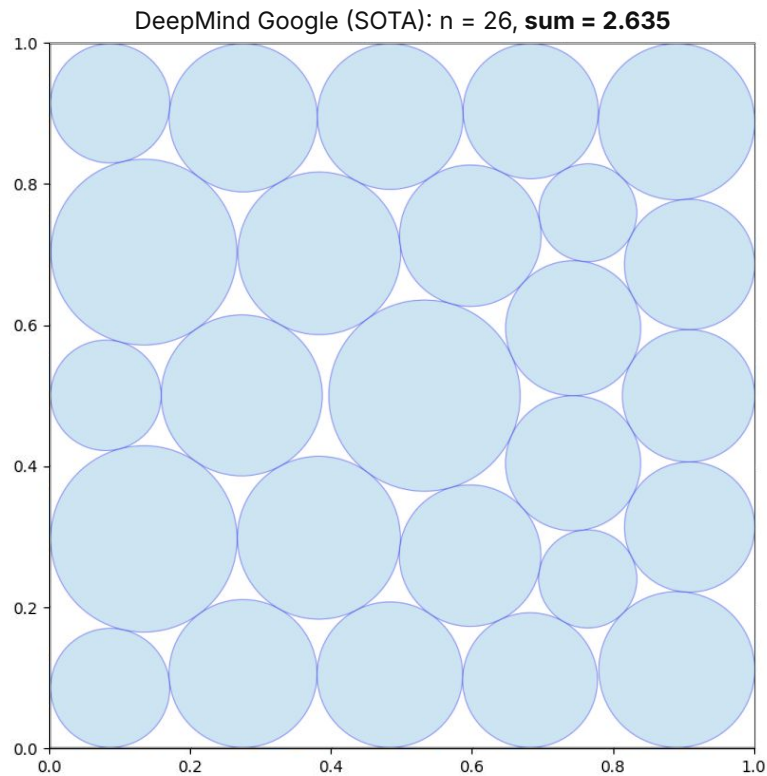
Фреймворк	Модели	Вес модели	Метрика
<b>MetaEvolve</b>	Qwen3 14B (thinking) Qwen2.5 Coder 14B	0.8 0.2	1.407
	Qwen3 14B (thinking) Qwen2.5 Coder 14B	0.2 0.8	1.468
	mistral-small mistral-medium	0.5 0.5	1.621
	Qwen3 14B (thinking) Qwen2.5 Coder 14B	0.5 0.5	1.682
	mistral-medium mistral-large	0.5 0.5	1.889
	mistral-small mistral-medium	0.8 0.2	2.322
	mistral-medium mistral-large	0.8 0.2	2.533
	<b>Qwen3 14B (thinking)</b> <b>Qwen2.5 Coder 14B</b>	<b>0.5</b> <b>0.5</b>	<b>2.624</b>

# Сравнение лучших результатов исследуемых реализаций AlphaEvolve с zero-shot режимом

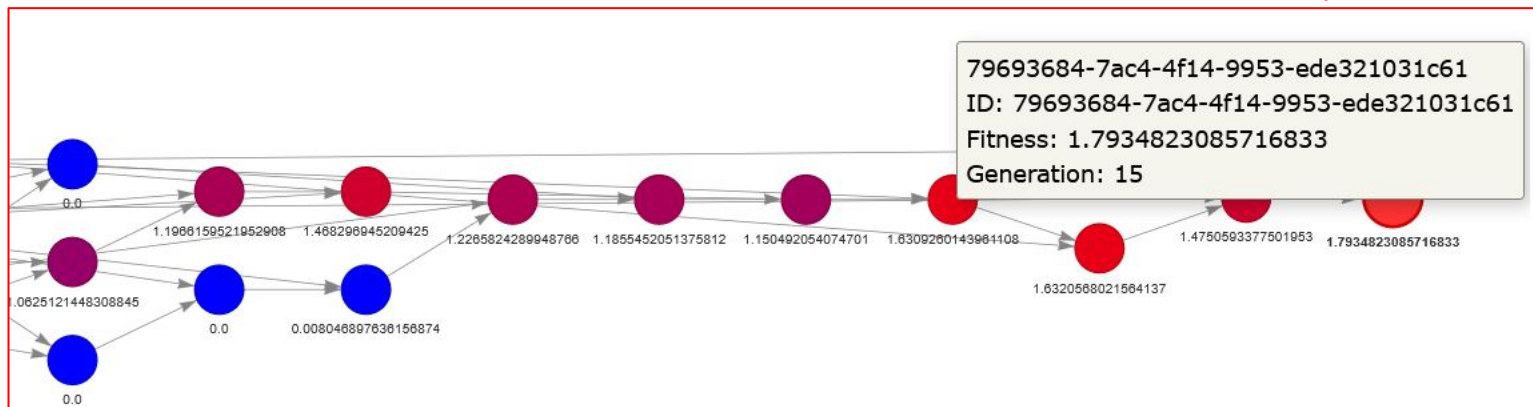
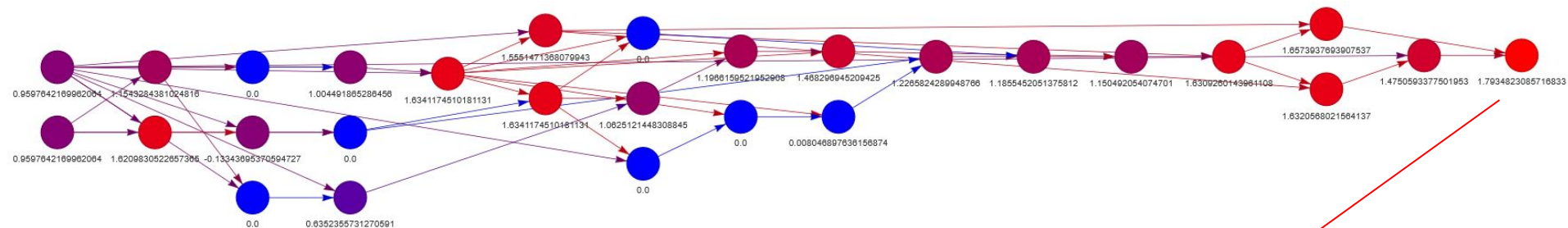
Фреймворк	Модели	Вес модели	Метрика
OpenEvolve	Qwen3 14B (thinking) Qwen2.5 Coder 14B	0.5 0.5	2.618
	deepseek-v3-0324 gemini-2.0-flash-exp	0.2 0.8	2.620
MetaEvolve	mistral-small mistral-medium	0.8 0.2	2.322
	mistral-medium mistral-large	0.8 0.2	2.533
	<b>Qwen3 14B (thinking)</b> <b>Qwen2.5 Coder 14B</b>	<b>0.5</b> <b>0.5</b>	<b>2.624</b>

Модель	Метрика
DeepSeek R1	1.05
Gemini 2.5 Flash	1.18
Claude Sonnet 4	1.42
GPT o3	1.44
Gemini 2.5 Pro	1.46
Claude Sonnet 4 Thinking	1.95
<b>GPT o3 Pro</b>	<b>2.03</b>

# Сравнение результатов с SOTA



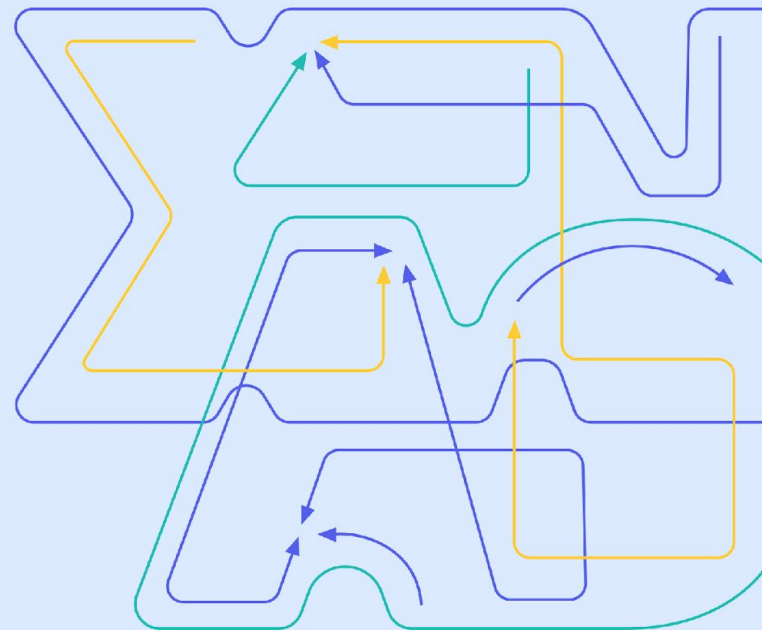
# Визуализация истории эволюции решений



# 05

---

Следующие шаги



# Следующие шаги

- 1 Анализ отслеживания инсайтов по древу эволюции
- 2 Агент мета-анализа популяции и ее истории для адаптивного управления процессом эволюции с опциональным режимом HITL
- 3 Улучшение эволюционного алгоритма
- 4 Адаптация под другие практически ориентированные задачи
- 5 Добавление других мультимодальных доменов для специфической задачи





# AIRI

## Команда



Денис Саматов  
@SamatovDS



Дмитрий Редько  
@dmitry\_redko



Елена Сухова  
@em\_suhova



Ануза Азеева  
@a-vestal



Виктор Волков  
@Rotverschiebung



Петр Анохин  
@petrkanokhin



Владимир Махарев  
@sm1rk



Данил Сивцов  
@svt\_danny