

# ИИ-агенты сегодня: обзор

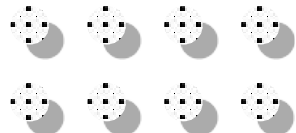
Путилова Елена  
2024 год

# О чем поговорим

- Кто такие ИИ-агенты, и куда мир катится
- Паттерны проектирования ИИ-агентов
- Строите свою GenAI-платформу? Тогда мы идем к вам!
- Что делать, если уже «чешутся лапки»?  
+ сразу небольшой дисклеймер



Illustrations by Pixeltrue on  
[icons8](#)

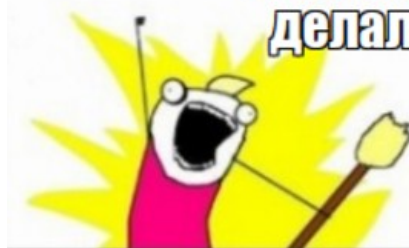


# Кто такие ИИ-агенты, и куда мир катится

- **Агенты — это система со сложными способностями к рассуждению, памятью и средствами для выполнения задач («Inroduction to LLM-agents», 2024, NVIDIA)**
- **Агентурные системы ИИ ... способностью предпринимать действия, которые ... способствуют достижению целей в течение длительного периода времени, без предварительного определения их поведения («Practices for governing agentic AI-systems», 2024, OpenAI)**
- **Искусственный агент... способности автономно планировать и выполнять последовательности действий, включая ... поиск информации, способом, способствующим достижению цели .., поставленной пользователем («The Ethics of Advanced AI Assistants», 2024, DeepMind)**



Мы хотим, чтобы ИИ  
делал все за нас!!!



Что  
именно?



Вообще  
все!!!



Когда мы это  
хотим?



Прямо  
сейчас!



# Переход к интерактивному ИИ

01

Традиционный ИИ

Решить аналитическую задачу

03

Интерактивный ИИ

Решить конечную задачу

02

Генеративный ИИ

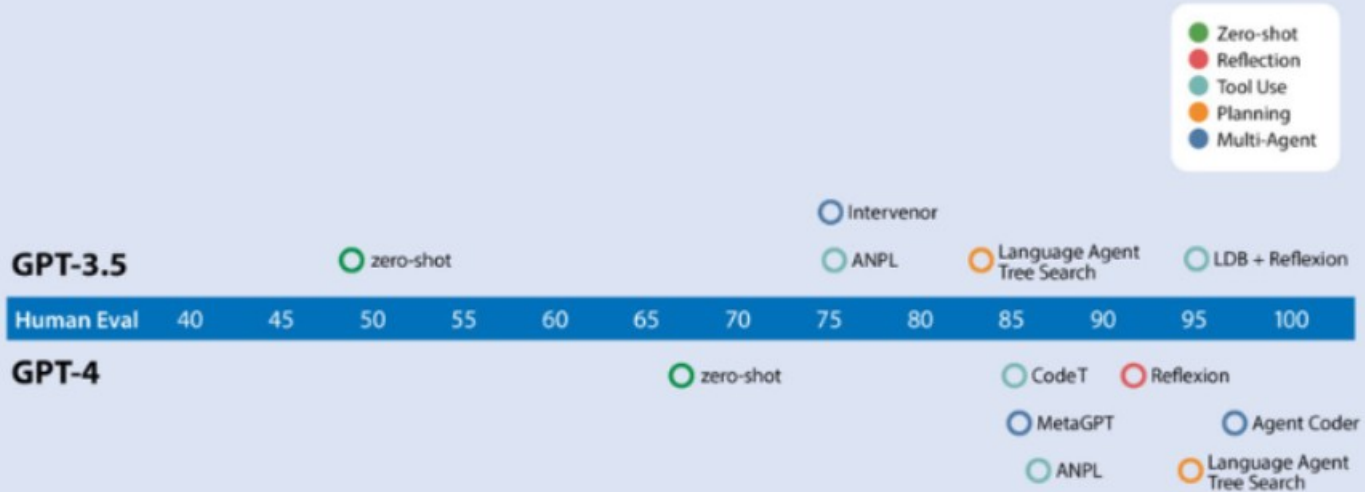
Создать контент

04 ?



# Зачем нужны ИИ-агенты

## GPT-3.5 and GPT-4 performance using zero-shot and agent workflows



Performance of GPT-3.5 and GPT-4 (zero-shot) on HumanEval, along with algorithms that use agent workflows on top of GPT-3.5 or GPT-4. Thanks to Joaquin Dominguez and John Santerre for help with this analysis.

<https://www.deeplearning.ai/the-batch/how-agents-can-improve-llm-performance/>

Эндрю Ын:

# Design Patterns: Towards Agency →

*Not an Agent*

*Single Agent*

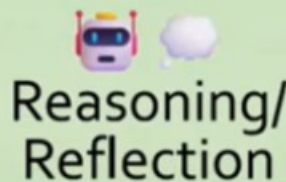
*Multi-Agent System*



LLM



Tool Use



Reasoning/  
Reflection



Planning/  
Execution



Multi-Agent  
Systems

*Robust Tech*

*Emerging Tech*

<https://www.deeplearning.ai/the-batch/how-agents-can-improve-llm-performance/>  
<https://www.youtube.com/live/FcjbkUcvQOs>

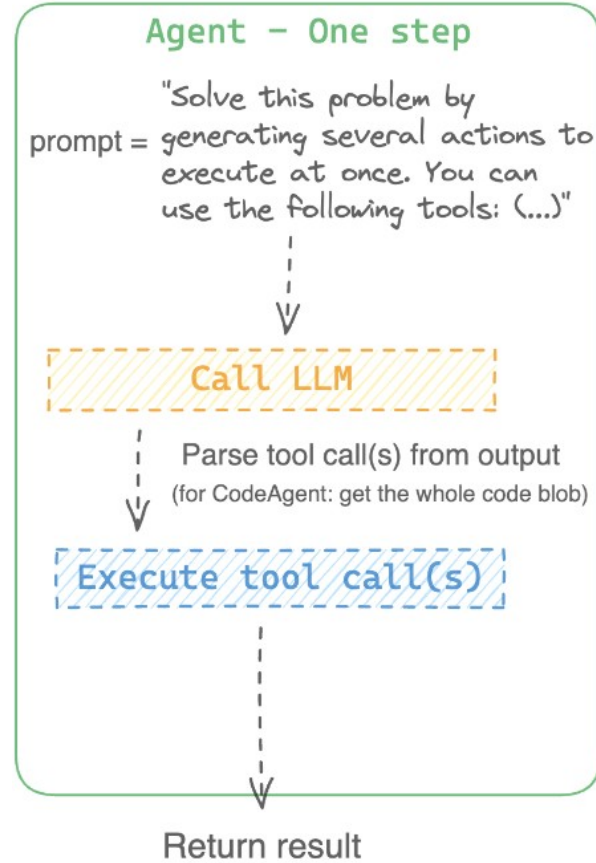
(с 42й минуты, выступление А. Попова, МФТИ)

## Tool use





## Tool use



<https://huggingface.co/docs/transformers/en/agents#what-is-an-agent>

# Reflection



# Agentic Design Patterns: Reflection



Please write code for {task}

```
def do_task(x): ...
```

```
def do_task_v2(x):
```

```
def do_task_v3(x):
```



Coder Agent  
(LLM)

There's a bug on line 5. Fix it by ...

It failed Unit Test 3. Try changing ...



Critic Agent  
(LLM)

## Tool use + Reflection = ReAct

### (1) Hotspot QA

**Question:** How many rooms are in the hotel that is home to the Cirque du Soleil show Mystere?

**HotspotQA label:** 2,664

Outdated X

### (1a) Standard

**Answer:** 3,000

X

### (1b) CoT (Reason Only)

**Thought:** Let's think step by step. The hotel that is home to the Cirque du Soleil show Mystere is Treasure Island. Treasure Island has 2,885 rooms, so the answer is 2,885.

**Answer:** 2,885

X

### (1c) Act-Only

**Act 1:** Search[Cirque du Soleil]  
**Obs 1:** Cirque du Soleil is a Canadian entertainment company...  
**Act 2:** Lookup[Mystere]  
**Obs 2:** No more results  
**Act 3:** Search[Mystere (Cirque du Soleil show)]  
**Obs 3:** Could not find...  
**Act 4:** Search[Mystere (Cirque du Soleil)]  
**Obs 4:** Mystere (my-steer) is one of six...  
**Act 5:** Lookup[Treasure Island Hotel and Casino]  
.....(end without answer)

X

### (1d) ReAct (Reason + Act)

**Thought 1:** I need to search Cirque du Soleil show Mystere, find the hotel it is in, then find the number of rooms in the hotel.

**Act 1:** Search[Cirque du Soleil show Mystere]

**Obs 1:** Could not find Cirque du Soleil show Mystere.

**Similar:** ['Cirque du Soleil' ...]

**Thought 2:** To find the hotel, I can search Mystere (Cirque du Soleil).

**Act 2:** Search[Mystere (Cirque du Soleil)]

**Obs 2:** Mystere (my-steer) is ... Held in a custom theatre at the Treasure Island Hotel and Casino ...

**Thought 3:** Mystere is held in a custom theatre at the Treasure Island Hotel and Casino. So I need to search Treasure Island Hotel and Casino next and find the number of rooms in the hotel.

**Act 3:** Search[Treasure Island Hotel and Casino]

**Obs 3:** Treasure Island Hotel and Casino ... with 2,884 rooms and 220 suites ...

**Thought 4:** Treasure Island Hotel and Casino has 2,884 rooms and 220 suites. So the answer is 3,104.

**Act 4:** Finish[3,104]

Up-to-date

✓

# Planning





## Chain-of-thought as planning

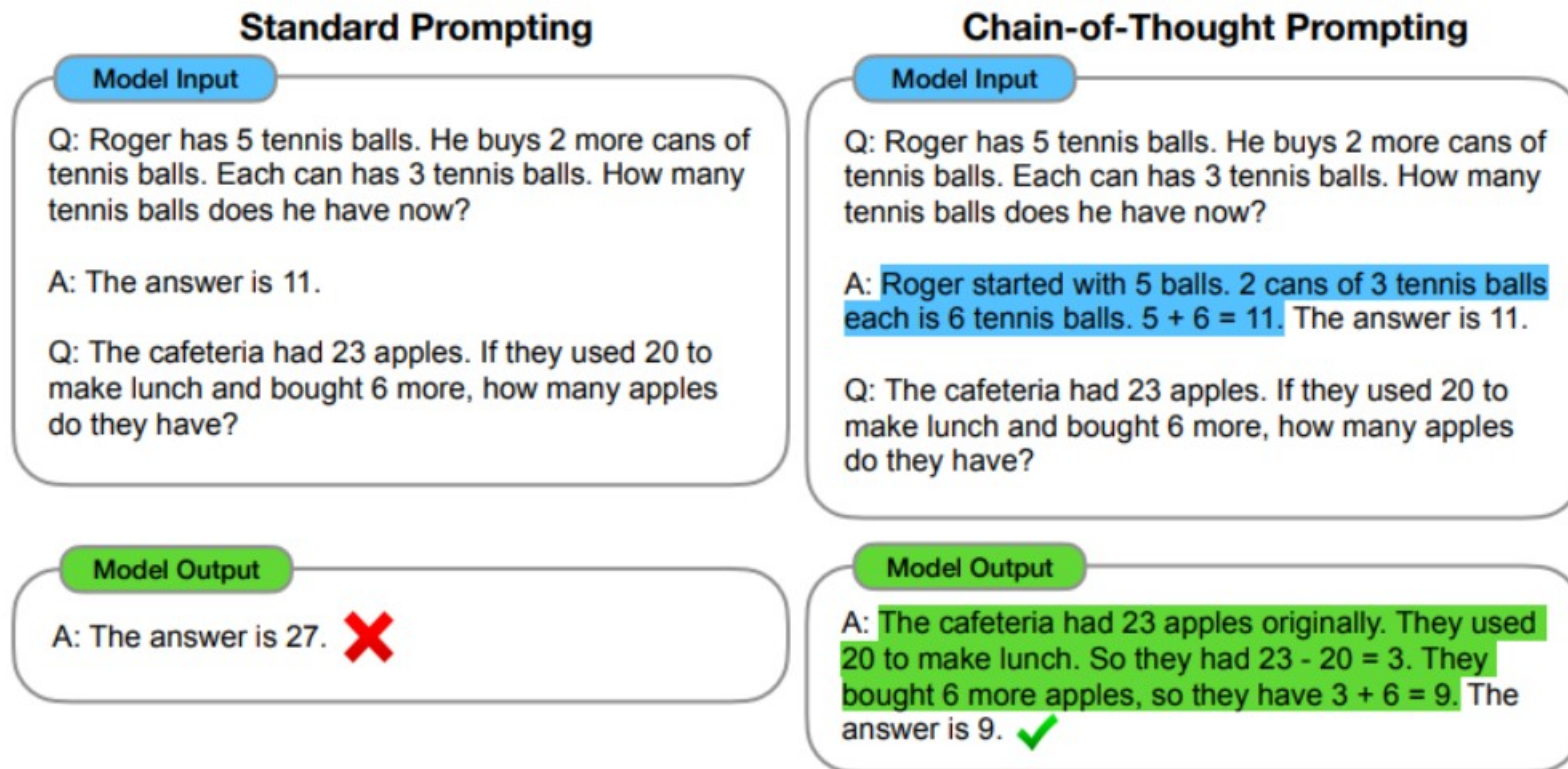


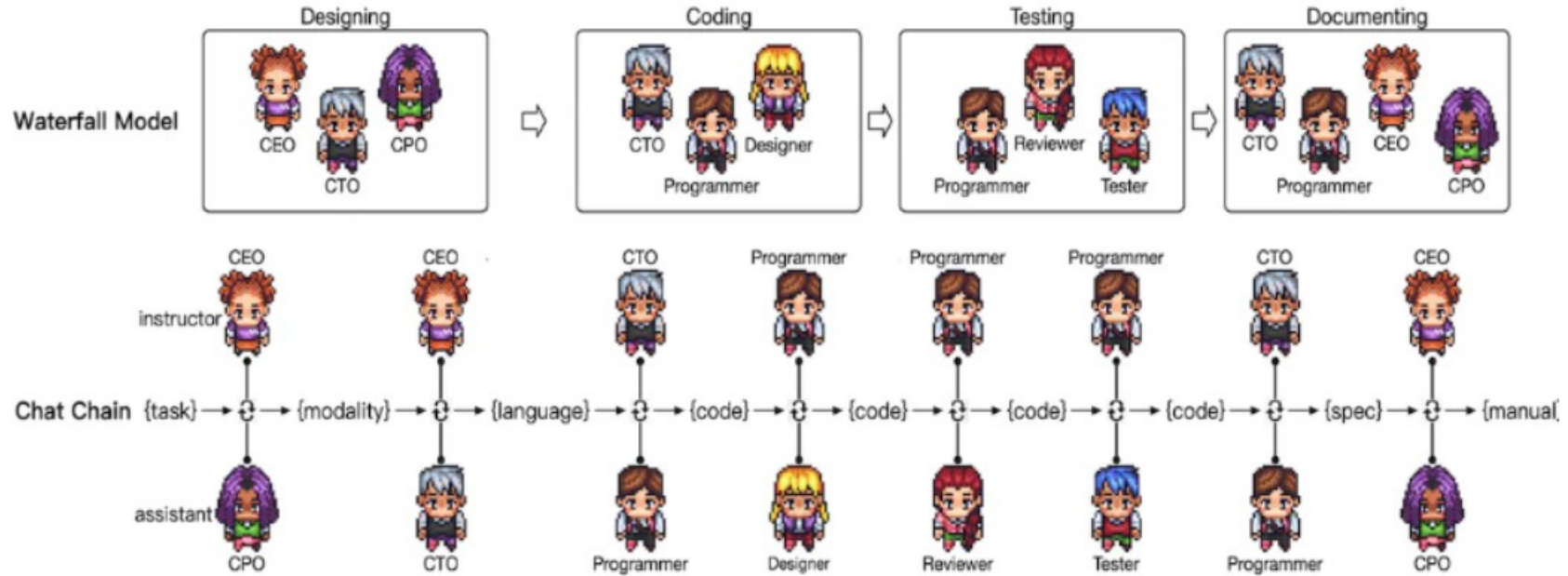
Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

# Multi-agent systems



# Multi-Agent Collaboration

## Agentic Design Patterns: Multi-Agent Collaboration



*Proposed ChatDev architecture. Image adapted from "Communicative Agents for Software Development," Qian et al. (2023).*



# Промежуточные выводы



Рефлексия и использование инструментов являются более понятными и надежными подходами



Планирование и коллаборация мультиагентов позволяет моделям делать больше, но это менее зрелые и менее предсказуемые подходы



Вероятно, что SLM на текущем уровне развития могут стать ИИ-агентами только на уровне tool use. Другие паттерны дизайна ИИ-агентов пока доступны только для SOTA среди LLM

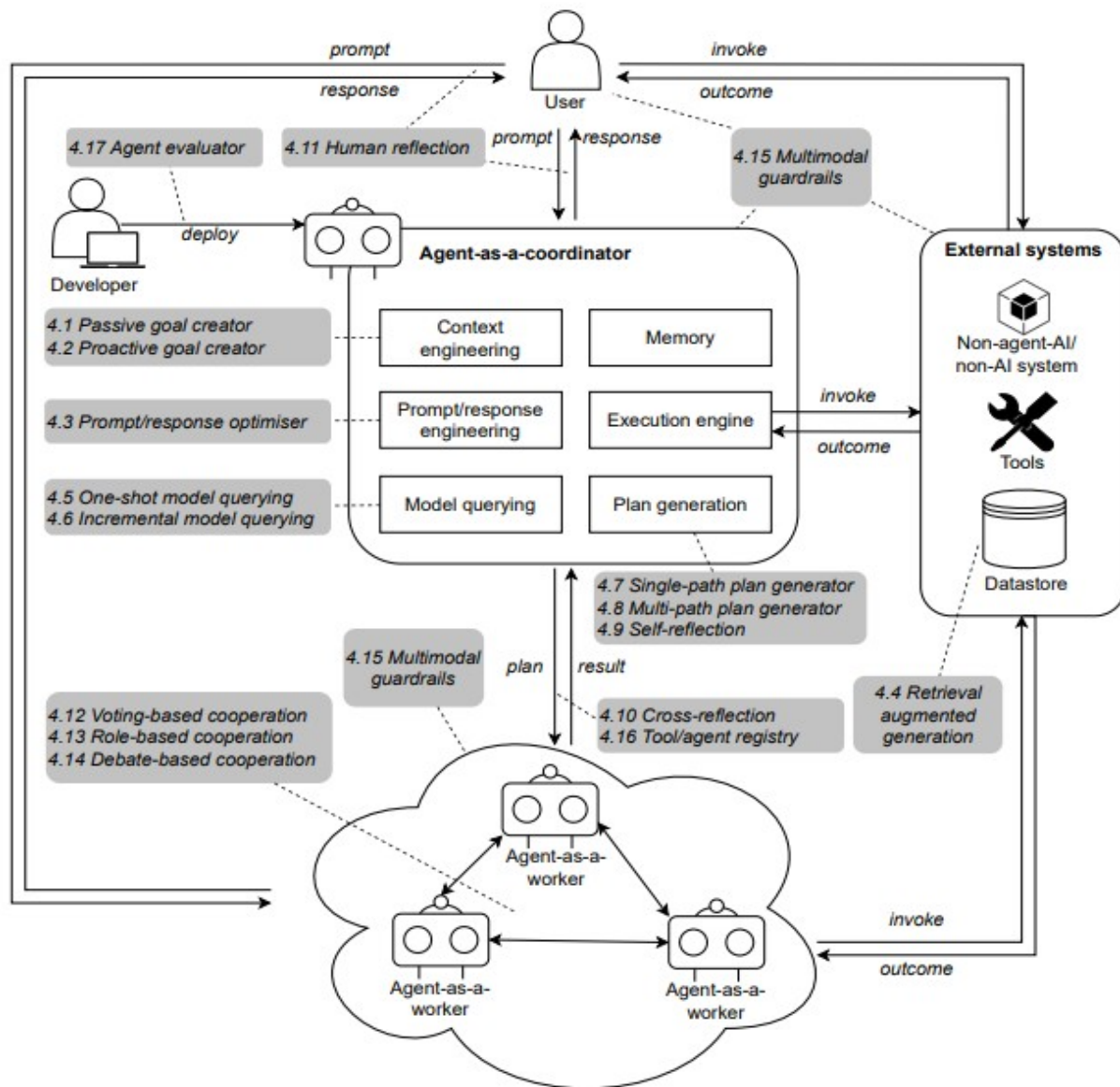
# AGENT DESIGN PATTERN CATALOGUE: A COLLECTION OF ARCHITECTURAL PATTERNS FOR FOUNDATION MODEL BASED AGENTS

Yue Liu, Sin Kit Lo, Qinghua Lu and others,  
<https://arxiv.org/html/2405.10467v1>



Экосистема  
различных паттернов  
ИИ-агентов (ИИ-  
агенты выделены  
серыми блоками)

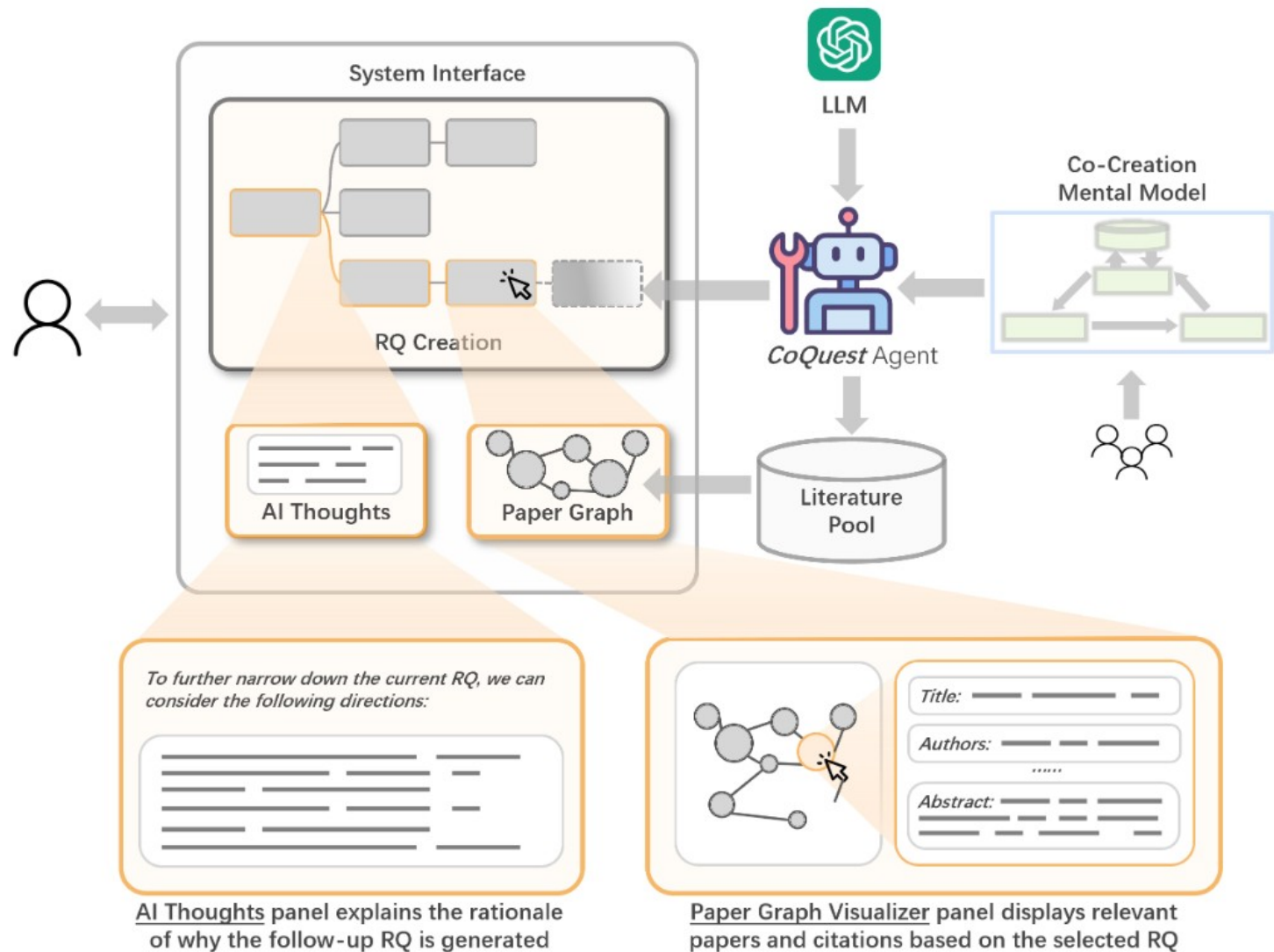
<https://arxiv.org/html/2405.10467v1>



Пассивный  
генератор целей.

Пример CoQuest:  
Exploring Research  
Question Co-Creation  
with an LLM-based  
Agent

[arXiv:2310.06155v3](https://arxiv.org/abs/2310.06155v3)



# Генератор плана. Пример

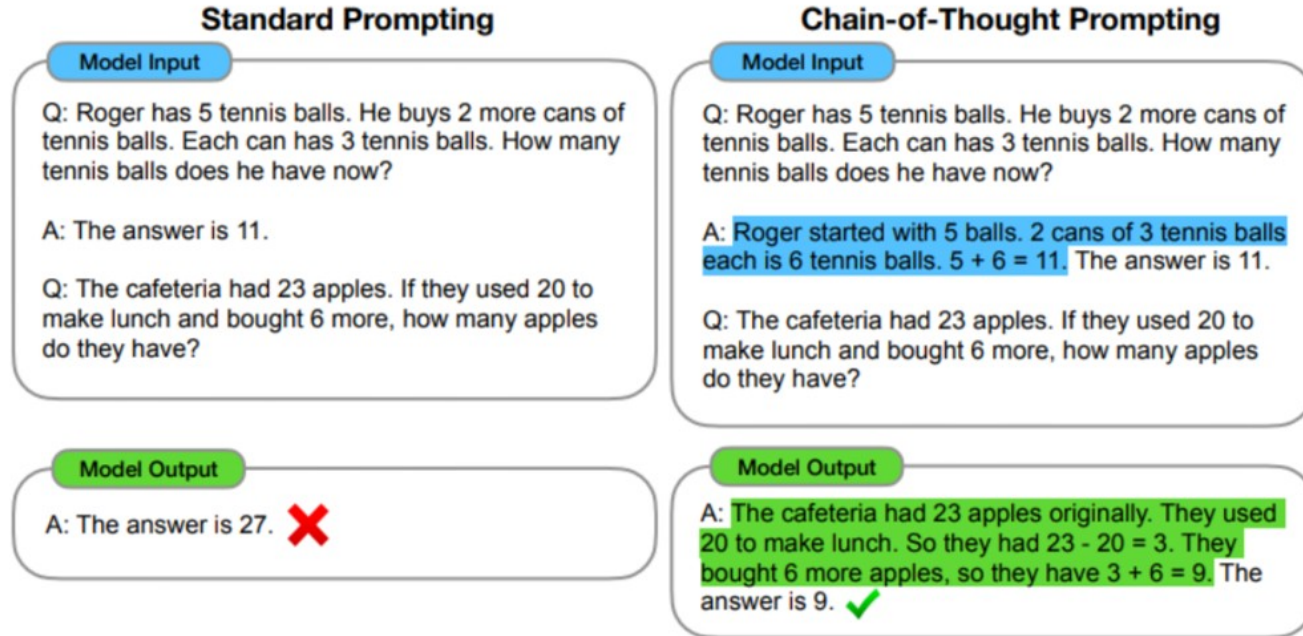


Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.



# Саморефлексия. Пример

Tool use + Reflection = ReAct

## (1) Hotspot QA

**Question:** How many rooms are in the hotel that is home to the Cirque du Soleil show Mystere?

HotspotQA label: 2,664

Outdated ✗

## (1a) Standard

Answer: 3,000

✗

## (1b) CoT (Reason Only)

**Thought:** Let's think step by step. The hotel that is home to the Cirque du Soleil show Mystere is Treasure Island. Treasure Island has 2,885 rooms, so the answer is 2,885.

Answer: 2,885

✗

## (1c) Act-Only

**Act 1:** Search[Cirque du Soleil]  
**Obs 1:** Cirque du Soleil is a Canadian entertainment company...  
**Act 2:** Lookup[Mystere]  
**Obs 2:** No more results  
**Act 3:** Search[Mystere (Cirque du Soleil show)]  
**Obs 3:** Could not find...  
**Act 4:** Search[Mystere (Cirque du Soleil)]  
**Obs 4:** Mystere (my-steer) is one of six...  
**Act 5:** Lookup[Treasure Island Hotel and Casino]  
.....(end without answer)

✗

## (1d) ReAct (Reason + Act)

**Thought 1:** I need to search Cirque du Soleil show Mystere, find the hotel it is in, then find the number of rooms in the hotel.

**Act 1:** Search[Cirque du Soleil show Mystere]

**Obs 1:** Could not find Cirque du Soleil show Mystere.

**Similar:** ['Cirque du Soleil' ...]

**Thought 2:** To find the hotel, I can search Mystere (Cirque du Soleil).

**Act 2:** Search[Mystere (Cirque du Soleil)]

**Obs 2:** Mystere (my-steer) is ... Held in a custom theatre at the Treasure Island Hotel and Casino ...

**Thought 3:** Mystere is held in a custom theatre at the Treasure Island Hotel and Casino. So I need to search Treasure Island Hotel and Casino next and find the number of rooms in the hotel.

**Act 3:** Search[Treasure Island Hotel and Casino]

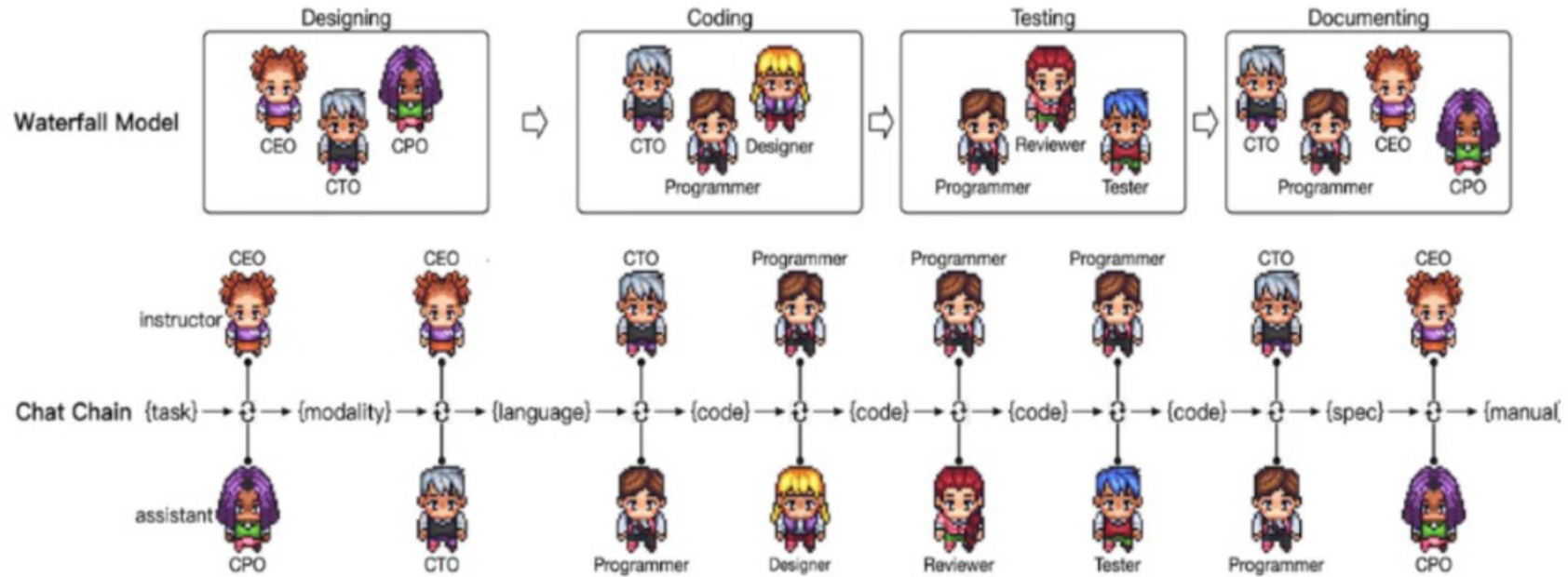
**Obs 3:** Treasure Island Hotel and Casino ... with 2,884 rooms and 220 suites ...

**Thought 4:** Treasure Island Hotel and Casino has 2,884 rooms and 220 suites. So the answer is 3,104.

**Act 4:** Finish[3,104]

Up-to-date ✓

# Кооперация агентов на основе ролей. Пример



Proposed ChatDev architecture. Image adapted from "Communicative Agents for Software Development," Qian et al. (2023).

<https://arxiv.org/abs/2307.07924>

## Реестр инструментов/агентов

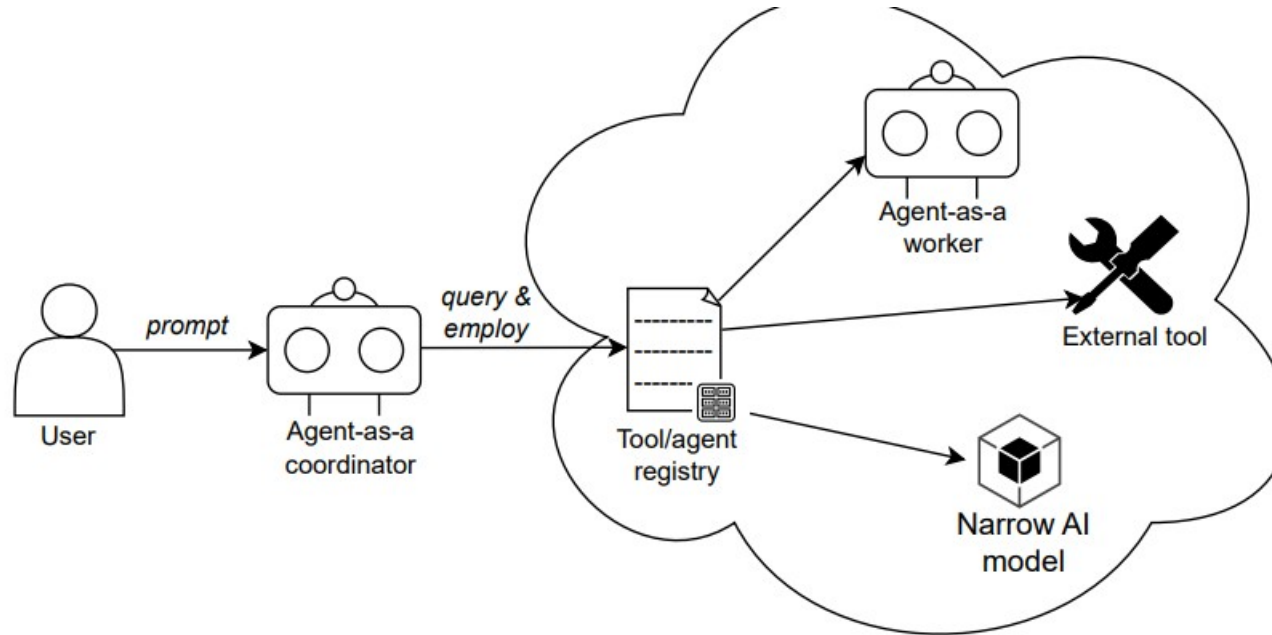
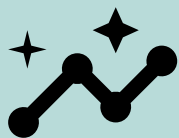


Figure 16: Tool/agent registry.



# Промежуточные выводы



Ценность исследования в  
широком обзоре



Также важно, что авторы  
предложили проект  
целой экосистемы ИИ-  
агентов



Данная классификация  
имеет некоторые  
пересечения с  
классификацией,  
предложенной Эндрю Ыном

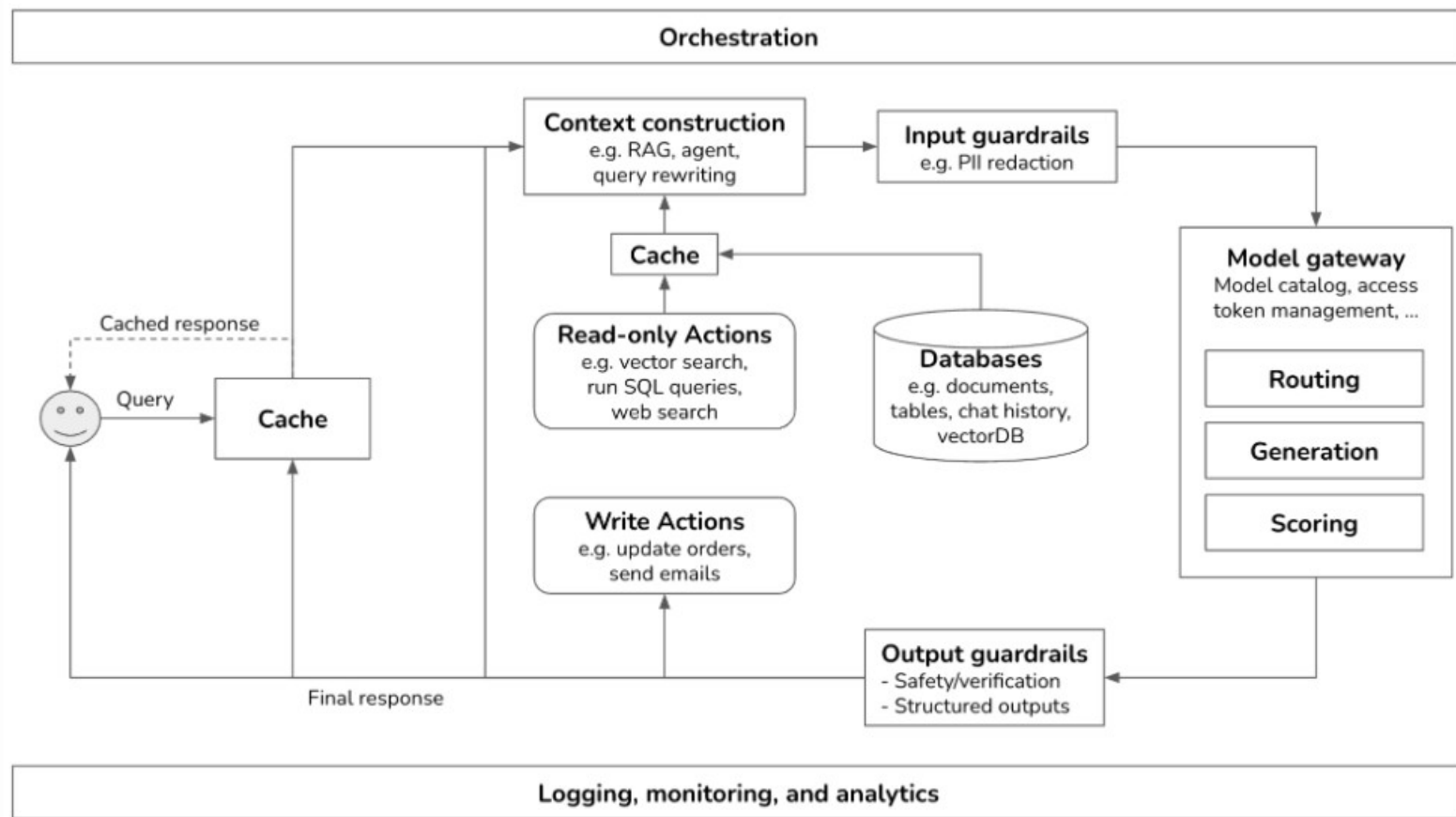
**Строите свою GenAI-  
платформу? Тогда мы  
идем к вам!**



МММ...  
ПИВО...

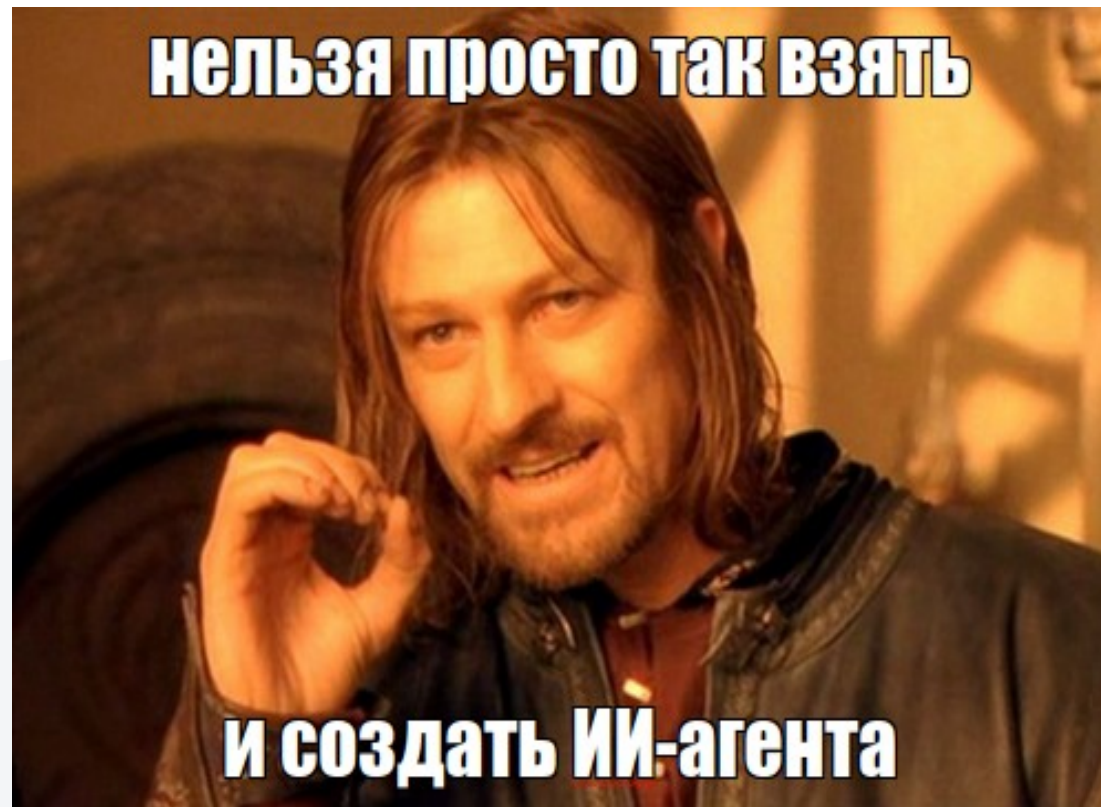
надо делать  
LLM-стартап  
на апишках

# Что нам советует Чип Хьюен:



## Сложности:

- Рассуждения, планирование
- Корректировка действий на основе обратной связи
- «Решения» мультиагентов
- Вопросы безопасности
- Вопросы этики



## Что делать, если уже «чешутся лапки»?

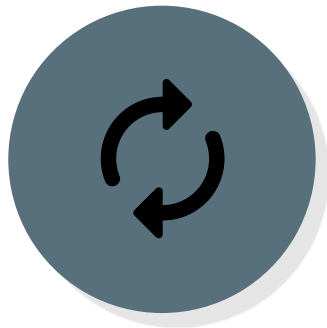
- DSPY (from Stanford University)  
Суперсила: умные промпты
- AutoGen (from Microsoft)  
Суперсила: удобно для мультиагентов
- LangGraph (from LangChain-team)  
Суперсила: план действий в графе



# Общие выводы



Обсудили, что принято  
включать в понятие ИИ-  
агентов



Рассмотрели какие  
паттерны  
проектирования ИИ-  
агентов можно выделить



Рассмотрели как можно  
использовать ИИ-агентов, а  
также какие фреймворки  
можно использовать, если  
вы строите свою GenAI-  
платформу

**Спасибо за внимание!**  
**Ваши вопросы**



Photo by [Dave Hoefler](#) on  
[Unsplash](#)