

# Введение в агенты с YandexGPT и Yandex Cloud

Дмитрий Сошников  
к.ф.-м.н., доцент МАИ и НИУ ВШЭ,  
консультант, Yandex Cloud

# Nvidia CEO Says 2025 Is the Year of AI Agents

By [Tae Kim](#) [Follow](#)

Jan 07, 2025, 5:40 pm EST

[Share](#) [AA](#) [Resize](#)

[Reprints](#) [Headphones](#)

[Nvidia](#) CEO Jensen Huang is optimistic that AI agents will become the next big thing for artificial intelligence.

## The AI agent market map

### Infrastructure

#### AI agent development platforms

[griptape](#) [A DEPT](#) [BREV/AN](#) [cohere](#) [DeepOpinion](#) [Distyl](#) [E2B](#)  
[Scma4.ai](#) [stack](#) [Kognitos](#) [kore.ai](#) [LangChain](#) [lyzr](#) [MISTRAL AI](#) [Moveworks](#)

#### Multi-agent & orchestration

[aaruu](#) [anuvai](#) [emergence](#) [fetch.ai](#) [Marian](#)

#### Authentication

[ANON](#) [Keet](#)

#### Web search & tool use

[Browserbase](#) [composio](#) [exa](#)  
[tavily](#)

#### Data curation

[LlamaIndex](#) [vectara](#)

#### Payments

[nevermined](#) [payman](#) [Skyfire](#)  
[stripe](#)

#### Memory

[Across AI](#)  
[Letta](#)  
[zep](#)

#### Evaluation & observability

[Agency](#) [arize](#) [COVAL](#)  
[Haize Labs](#) [Langfuse](#) [Patronus AI](#)  
[vijil](#) [Vocera](#) [Weights & Biases](#)

#### Voice

[Bland](#) [Cartesia](#) [Deepgram](#) [ElevenLabs](#)  
[FIXIE](#) [Hamming](#) [Retell AI](#) [sarvam.ai](#)  
[Synthflow](#) [VAPI](#) [WaveForms AI](#)

### Horizontal applications & job functions

#### Productivity & personal assistants

[ANTHROPIC](#) [Convergence](#) [HyperWrite](#) [OpenAI](#)  
[perplexity](#) [Personal AI](#) [Please](#) [BROWSER COMPANY](#)

#### General enterprise workflows

[AISERA](#) [Azara](#) [bardeen](#) [Ema](#)  
[glean](#) [HERCULESAI](#) [Orby](#) [Relevance AI](#)  
[RHINO AI](#) [Sana](#) [Senso](#) [SuperAGI](#)

#### Customer service

[COGNIGY](#) [crescendo](#) [Decagon](#)  
[DevRev](#) [Gradient Labs](#) [Maven AGI](#)  
[parloa](#) [SIERRA](#) [Voiceflow](#)  
[yampa](#)

#### Software development

[All Hands](#) [Autify](#) [Cognition](#) [CURSOR](#)  
[FACTORY](#) [imbue](#) [lovable](#) [Magic](#)  
[poolside](#) [qodo](#) [replit](#) [Resolve.ai](#)  
[Tusk](#)

#### Data analysis

[Athena Intelligence](#)  
[Connecty AI](#)  
[Paradigm](#)

#### Accounting

[Basis](#)  
[Hypatos](#)

#### Cybersecurity

[AIRMDR](#) [Bricklayer AI](#) [Dropzone AI](#)  
[nullify](#) [prophet](#) [RadiantSecurity](#)  
[7AI](#) [Simbian](#) [torq](#)

#### Sales

[11x](#) [AISDR](#) [ARTISAN](#) [bounti](#)  
[clay](#) [DocketAI](#) [flowworks](#) [mindy](#)  
[QUALIFIED](#) [IOX](#) [tektonic ai](#)

#### HR

[apriora](#) [Borderless AI](#) [CONVERZAI](#) [MERCOR](#)  
[moonhub](#) [otto](#) [Tezi](#)

#### Marketing

[ability.ai](#)  
[firsthand](#)

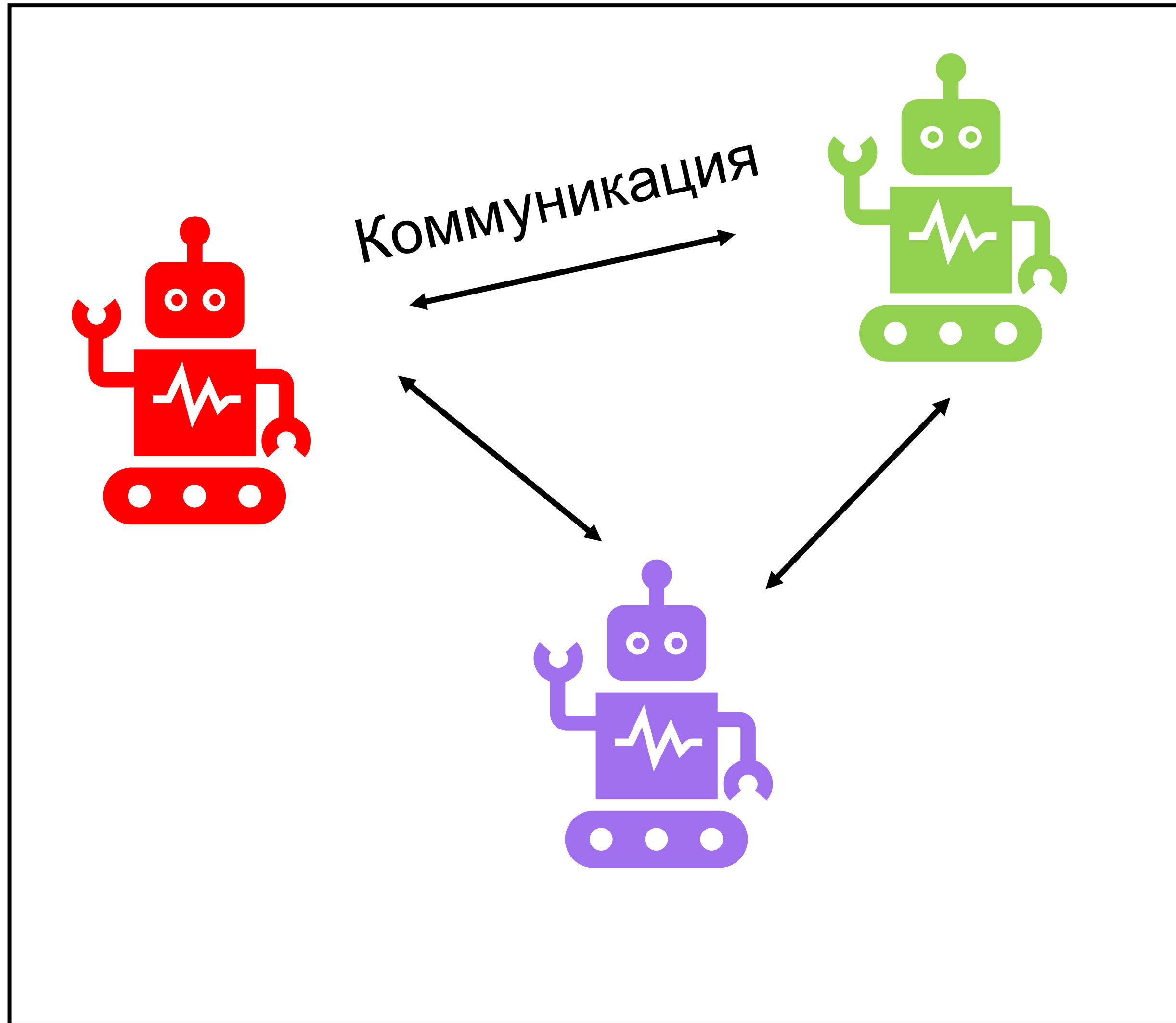
#### Web research & data extraction

[Lutra](#) [Reworkd](#) [XI](#)



# 199x

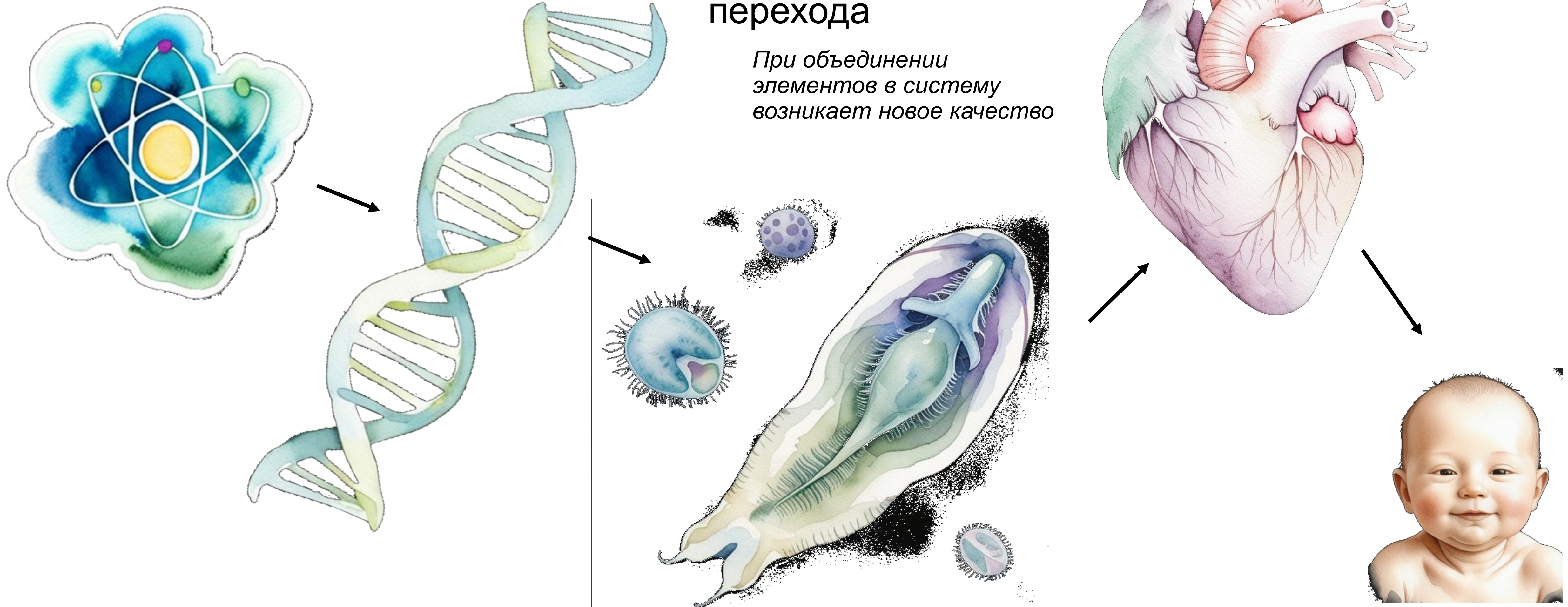
## Окружение



# Мотивация 1: Эволюционная кибернетика

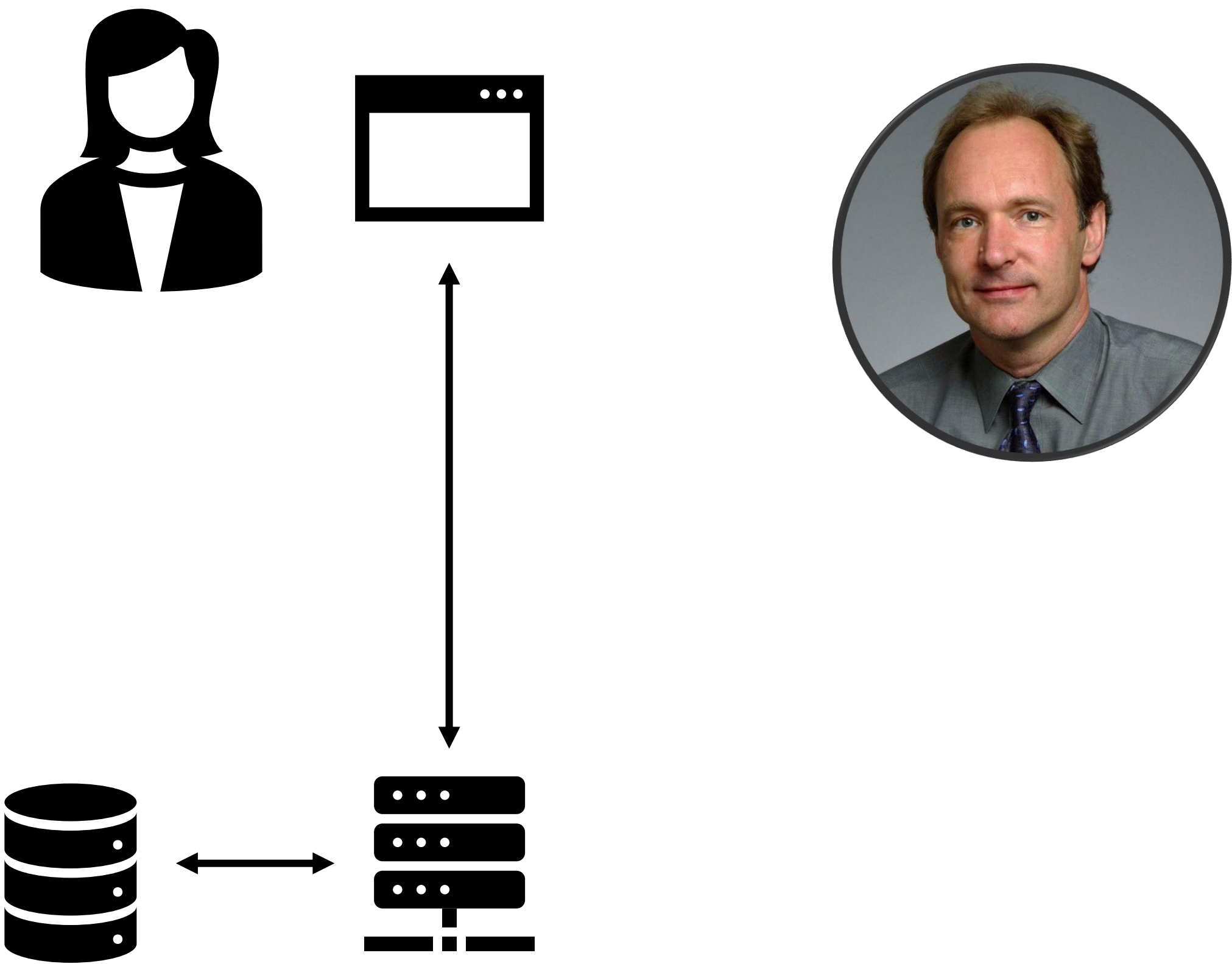
Принцип  
метасистемного  
перехода

*При объединении  
элементов в систему  
возникает новое качество*

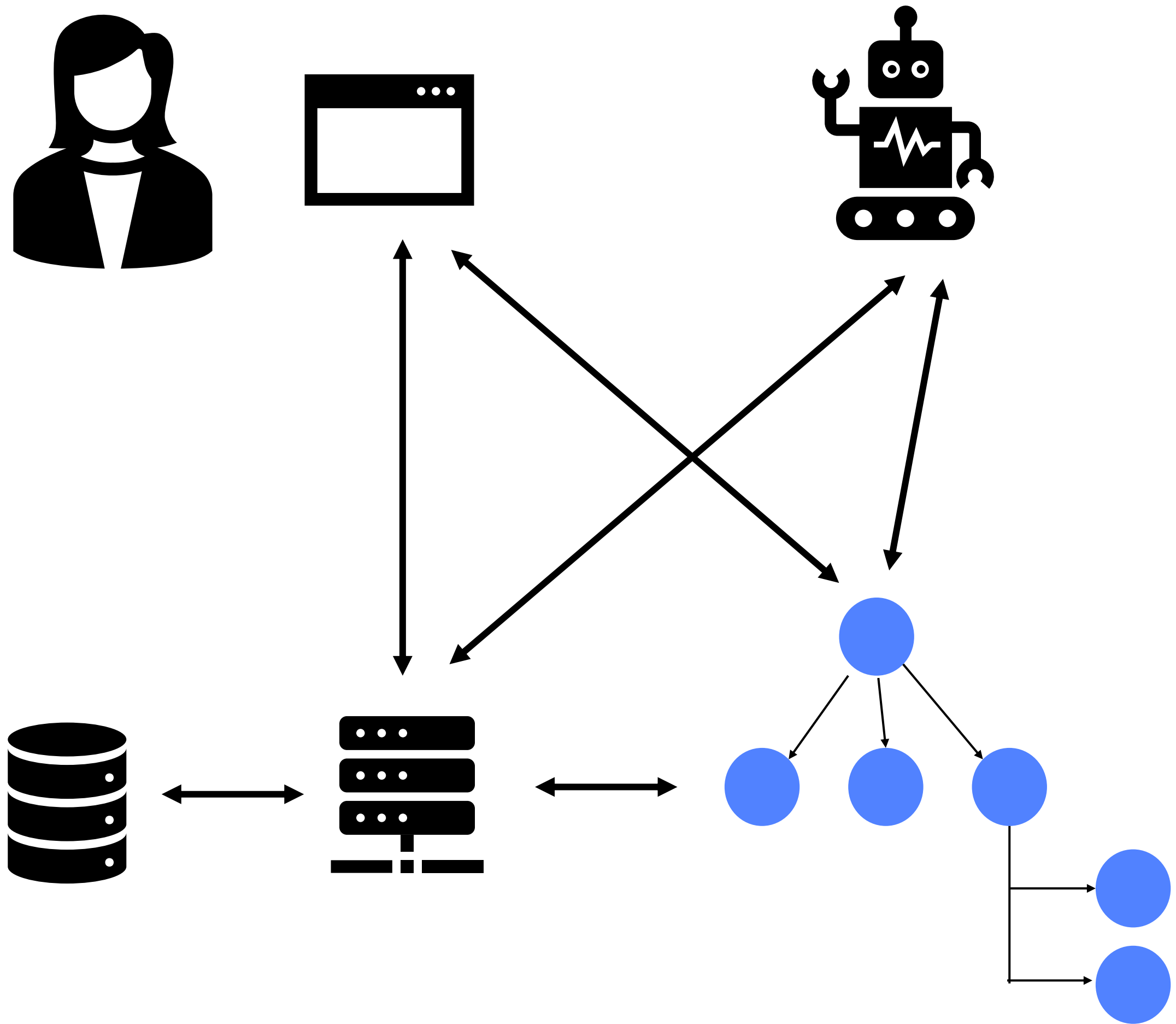


# Мотивация 2

Web



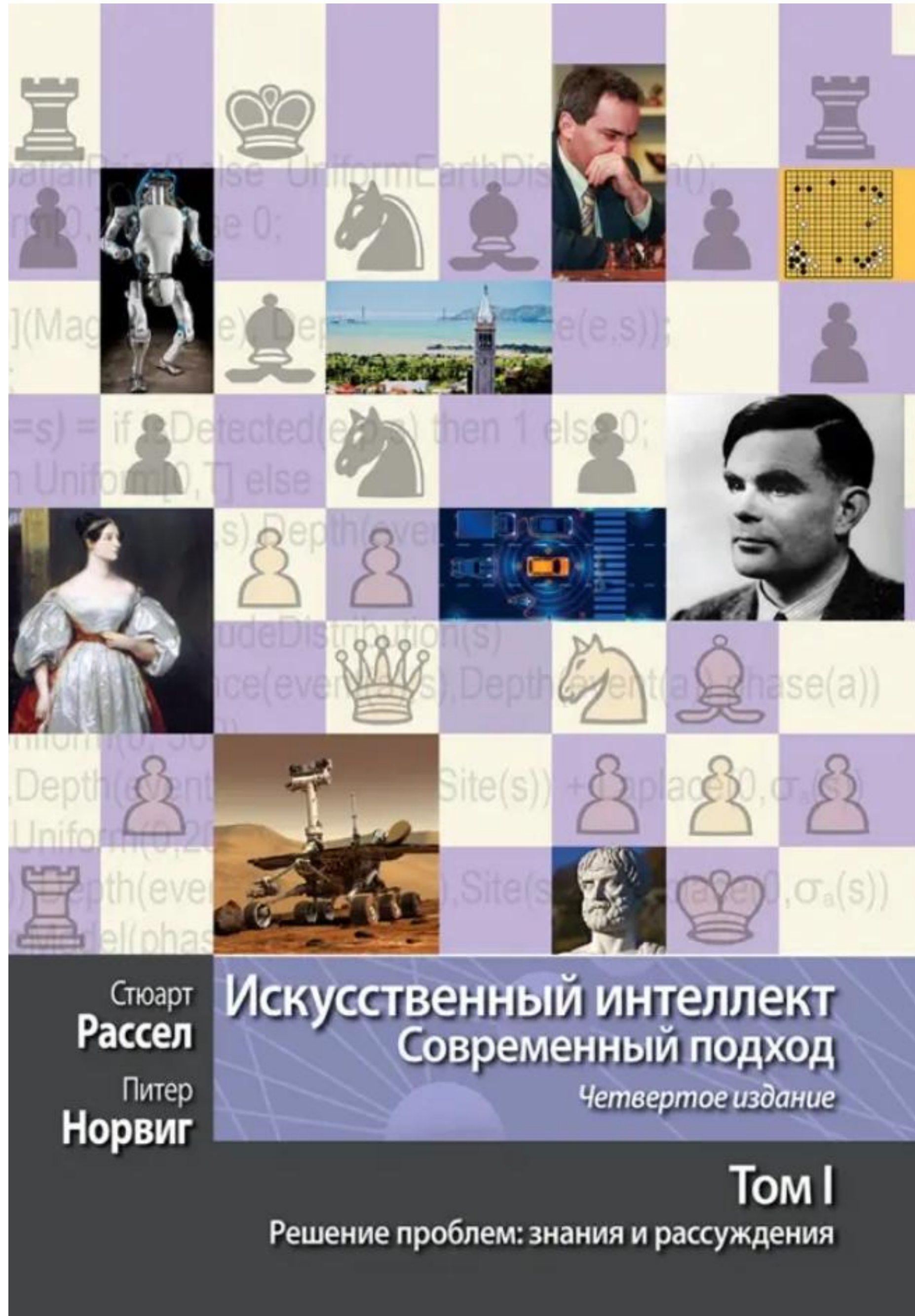
Semantic Web



# Мотивация 3



# 1995



[Preface \(pdf\)](#); [Contents with subsections \(pdf\)](#).

## **I Artificial Intelligence**

1 Introduction

2 Intelligent Agents

## **II Problem-solving**

3 Solving Problems by Searching

4 Search in Complex Environments

5 Constraint Satisfaction Problems

6 Adversarial Search and Games

## **III Knowledge, reasoning, and planning**

7 Logical Agents

8 First-Order Logic

9 Inference in First-Order Logic

10 Knowledge Representation

11 Automated Planning

## **IV Uncertain knowledge and reasoning**

12 Quantifying Uncertainty

13 Probabilistic Reasoning

14 Probabilistic Reasoning over Time

15 Making Simple Decisions

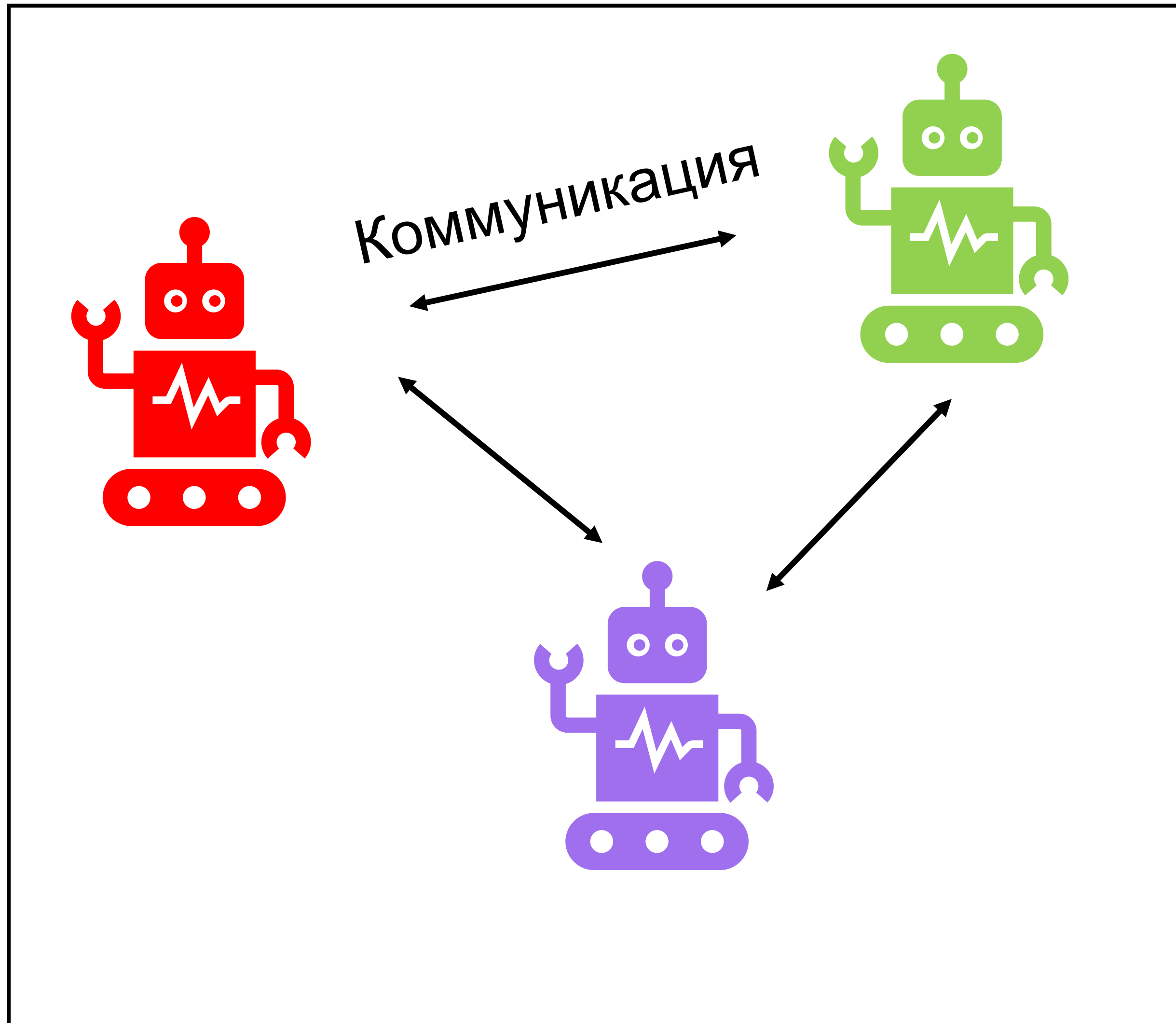
16 Making Complex Decisions

17 Multiagent Decision Making

18 Probabilistic Programming

# 199x

## Окружение



Автономность

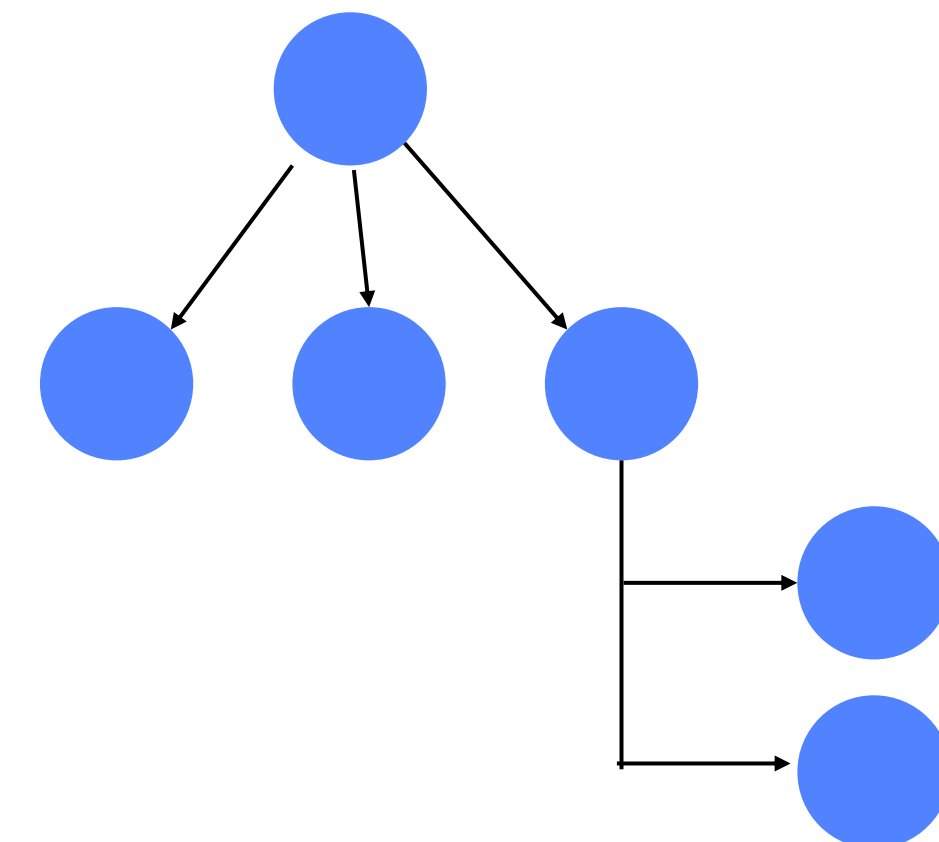
Реактивность

Проактивность

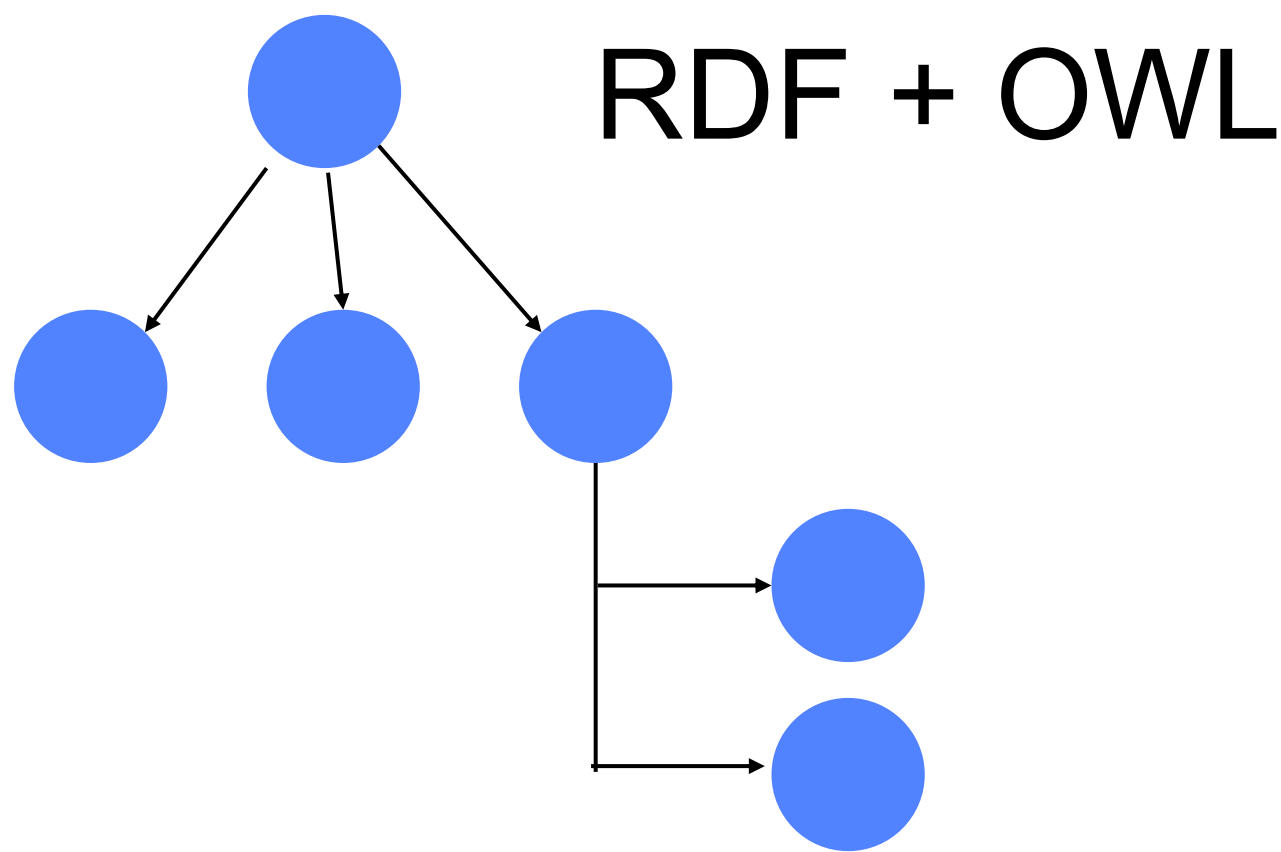
Социальность

Коммуникация

- Теория коммуникативных актов
- Теория аукционов
- Общие знаний (онтологии)



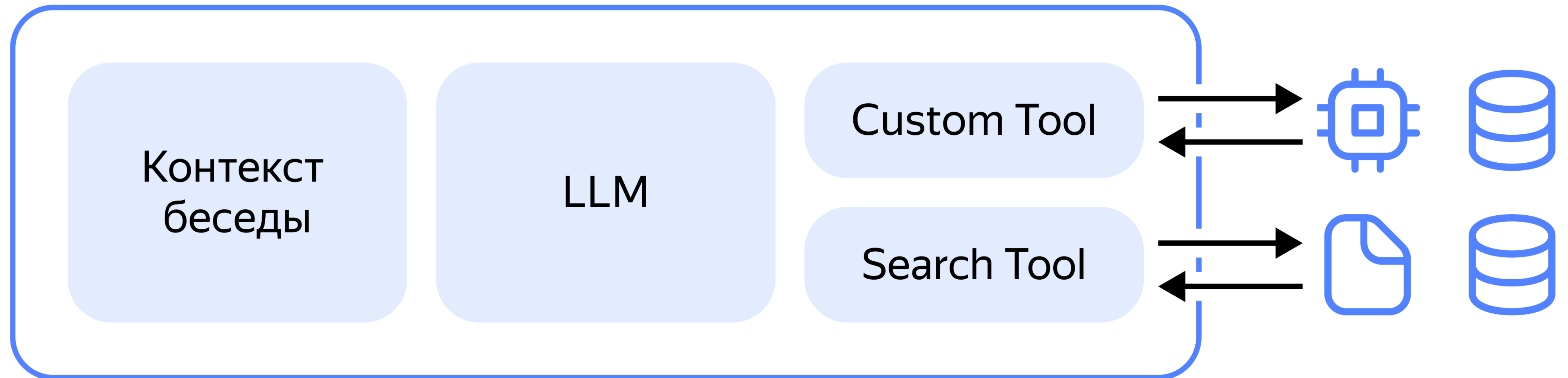
# Многоагентные системы в 1990-х годах



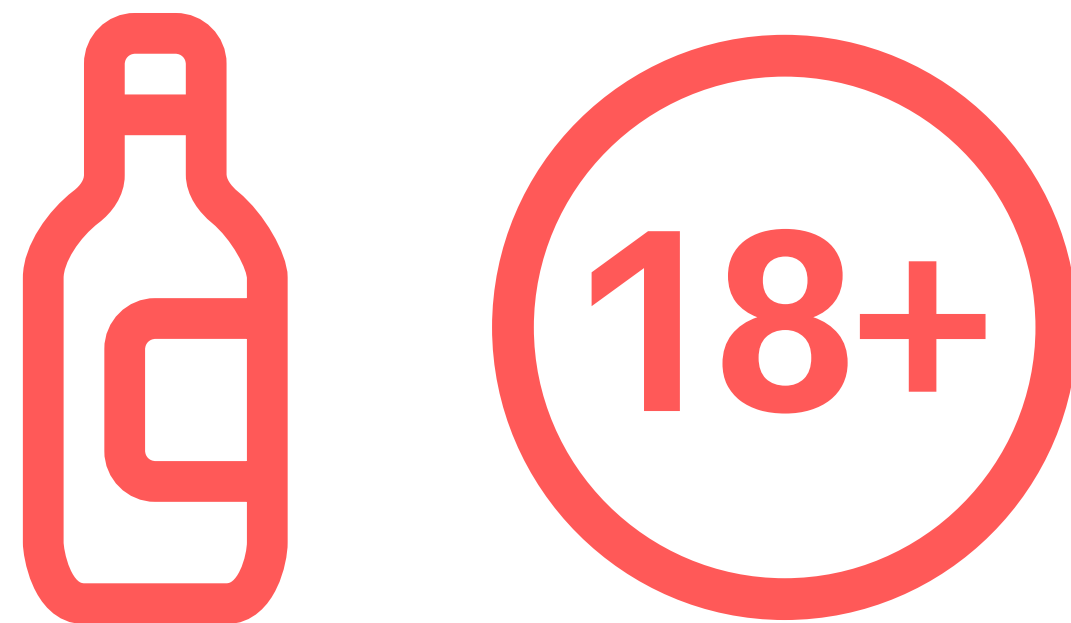
# Многоагентные системы в 2024-25



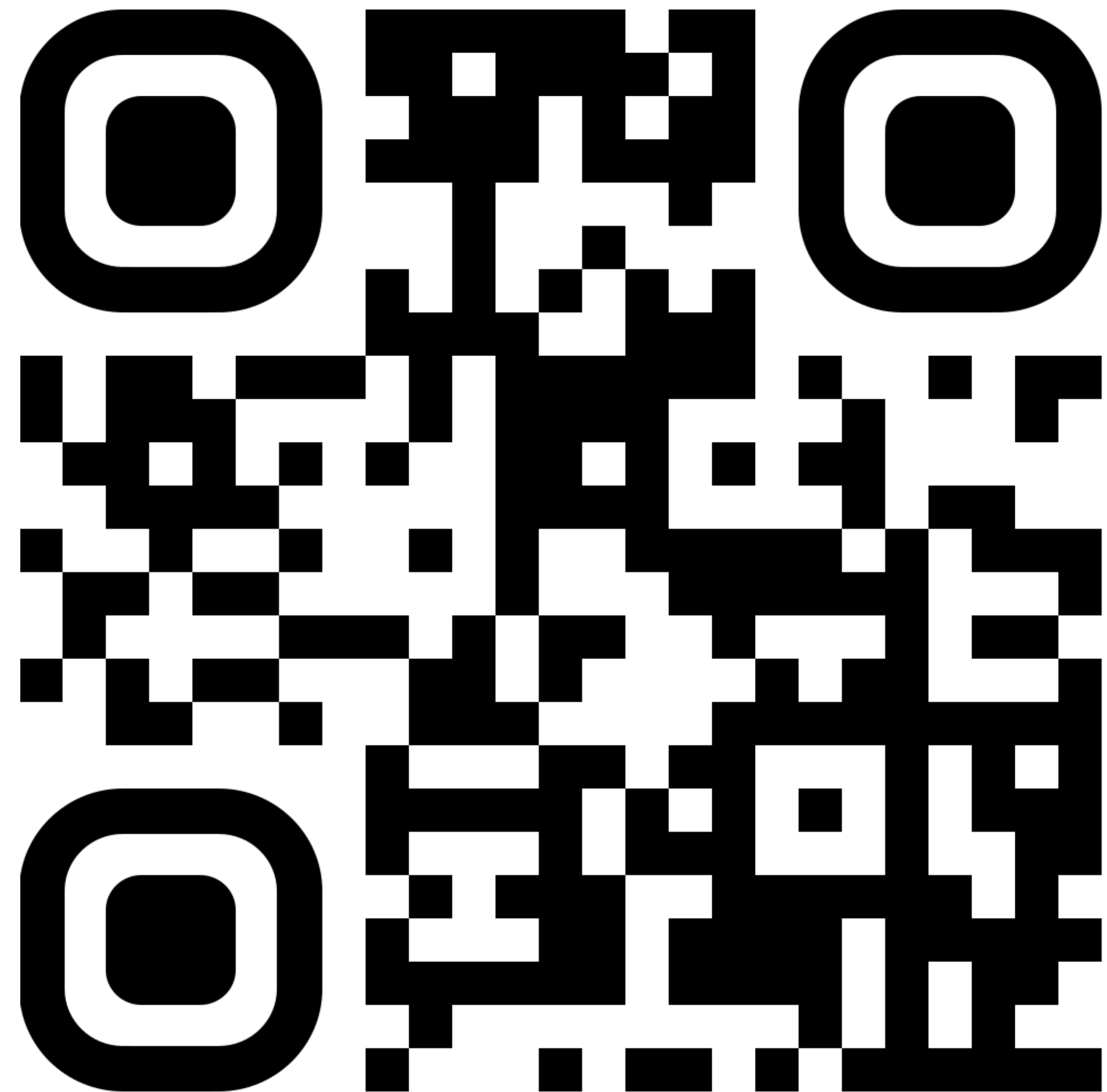
# Простейший агент = Ассистент



# Демонстрация

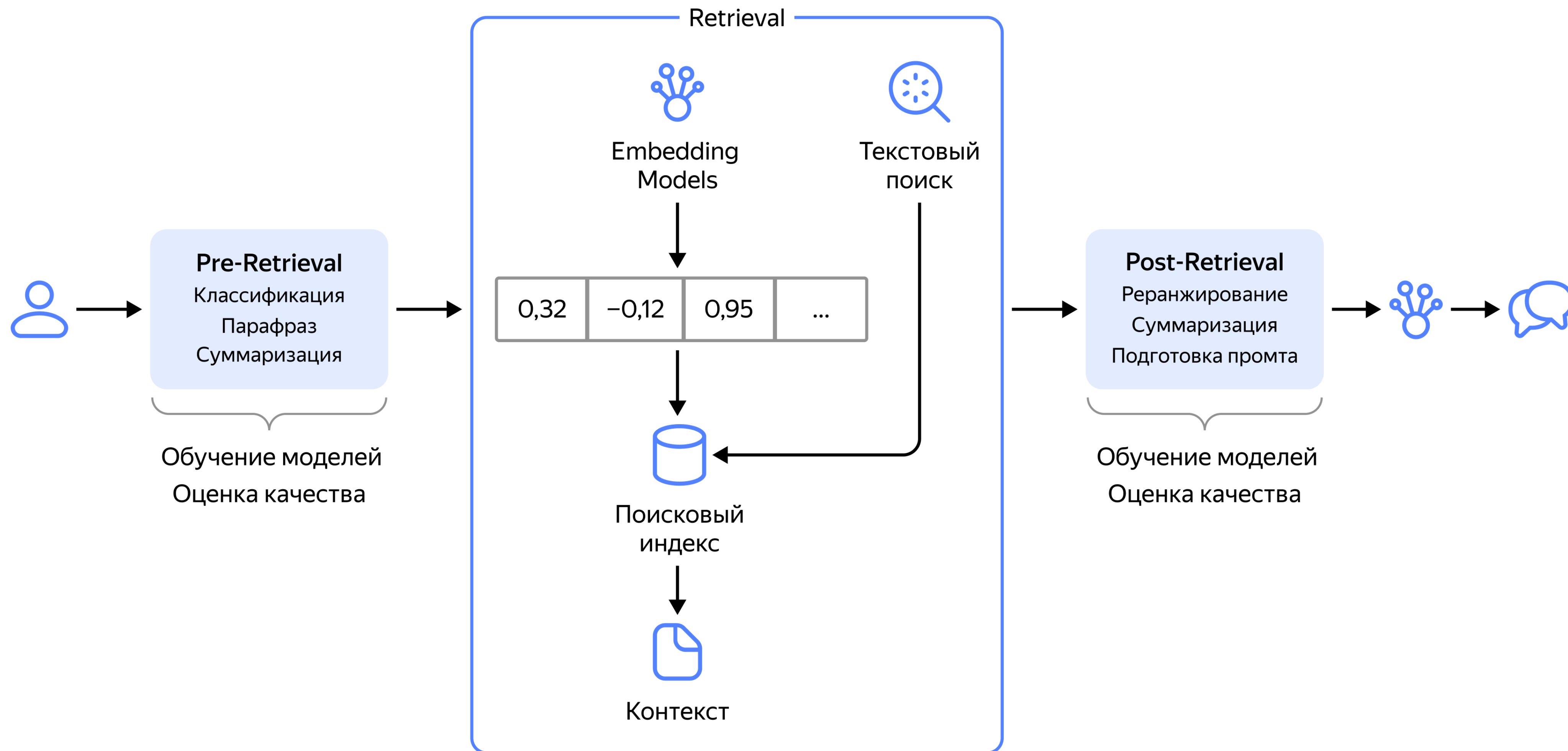


Употребление алкоголя  
вредит вашему здоровью...



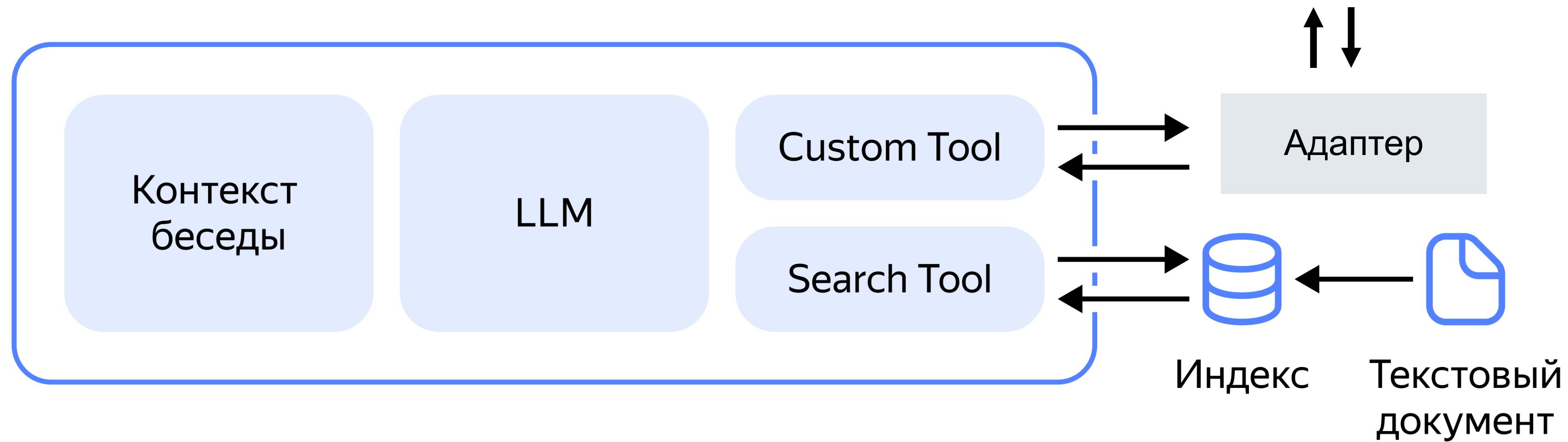
[github.com/yandex-datasphere/  
advanced-assistant](https://github.com/yandex-datasphere/advanced-assistant)

# Retrieval and Generation Workflow

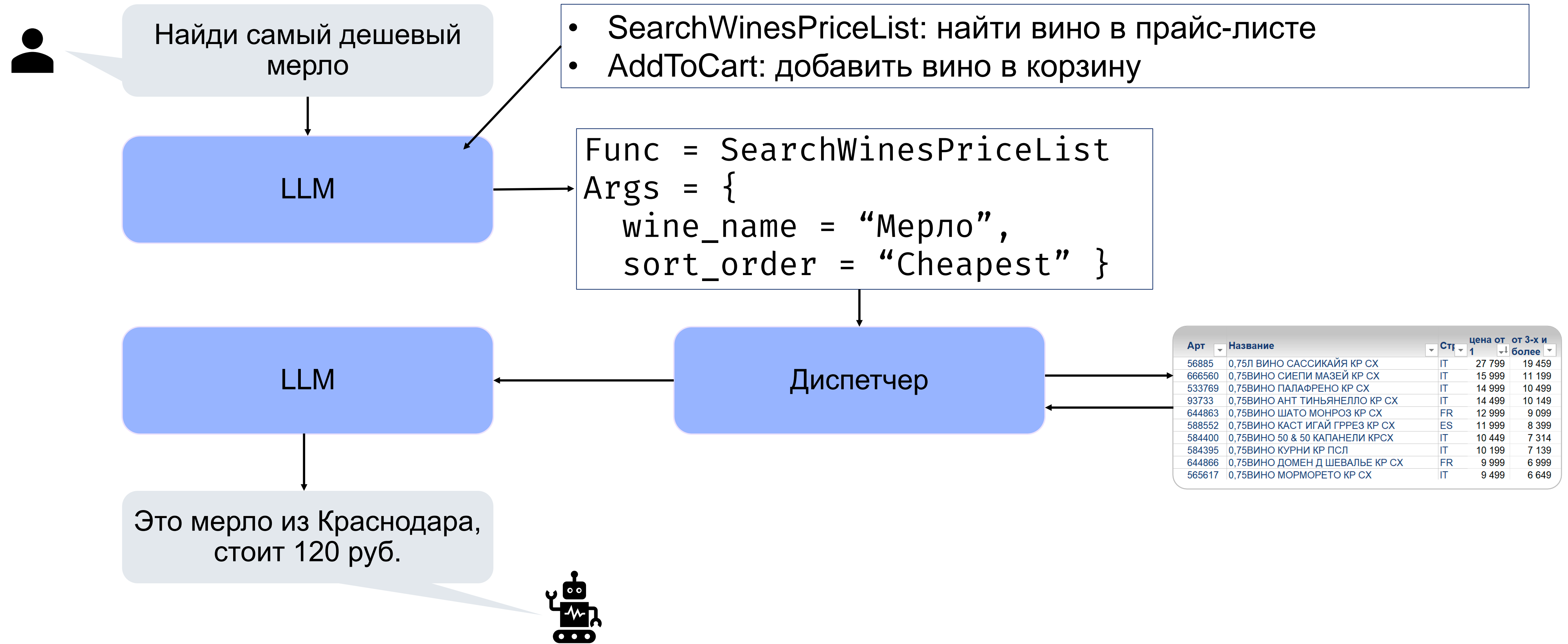


# Не всё ложится в RAG

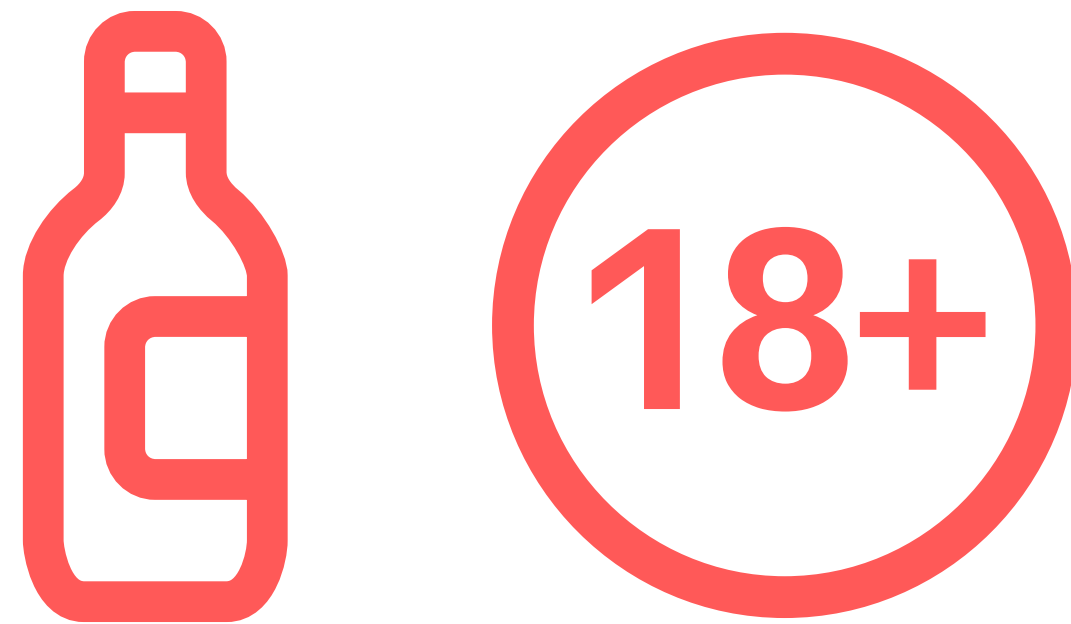
Арт	Название	Стр	цена от 1	от 3-х и более
56885	0,75Л ВИНО САССИКАЙЯ КР СХ	IT	27 799	19 459
666560	0,75ВИНО СИЕПИ МАЗЕЙ КР СХ	IT	15 999	11 199
533769	0,75ВИНО ПАЛАФРЕНО КР СХ	IT	14 999	10 499
93733	0,75ВИНО АНТ ТИНЯНЕЛЛО КР СХ	IT	14 499	10 149
644863	0,75ВИНО ШАТО МОНРОЗ КР СХ	FR	12 999	9 099
588552	0,75ВИНО КАСТ ИГАЙ ГРРЕЗ КР СХ	ES	11 999	8 399
584400	0,75ВИНО 50 & 50 КАПАНЕЛИ КРСХ	IT	10 449	7 314
584395	0,75ВИНО КУРНИ КР ПСЛ	IT	10 199	7 139
644866	0,75ВИНО ДОМЕН Д ШЕВАЛЬЕ КР СХ	FR	9 999	6 999
565617	0,75ВИНО МОРМОРЕТО КР СХ	IT	9 499	6 649



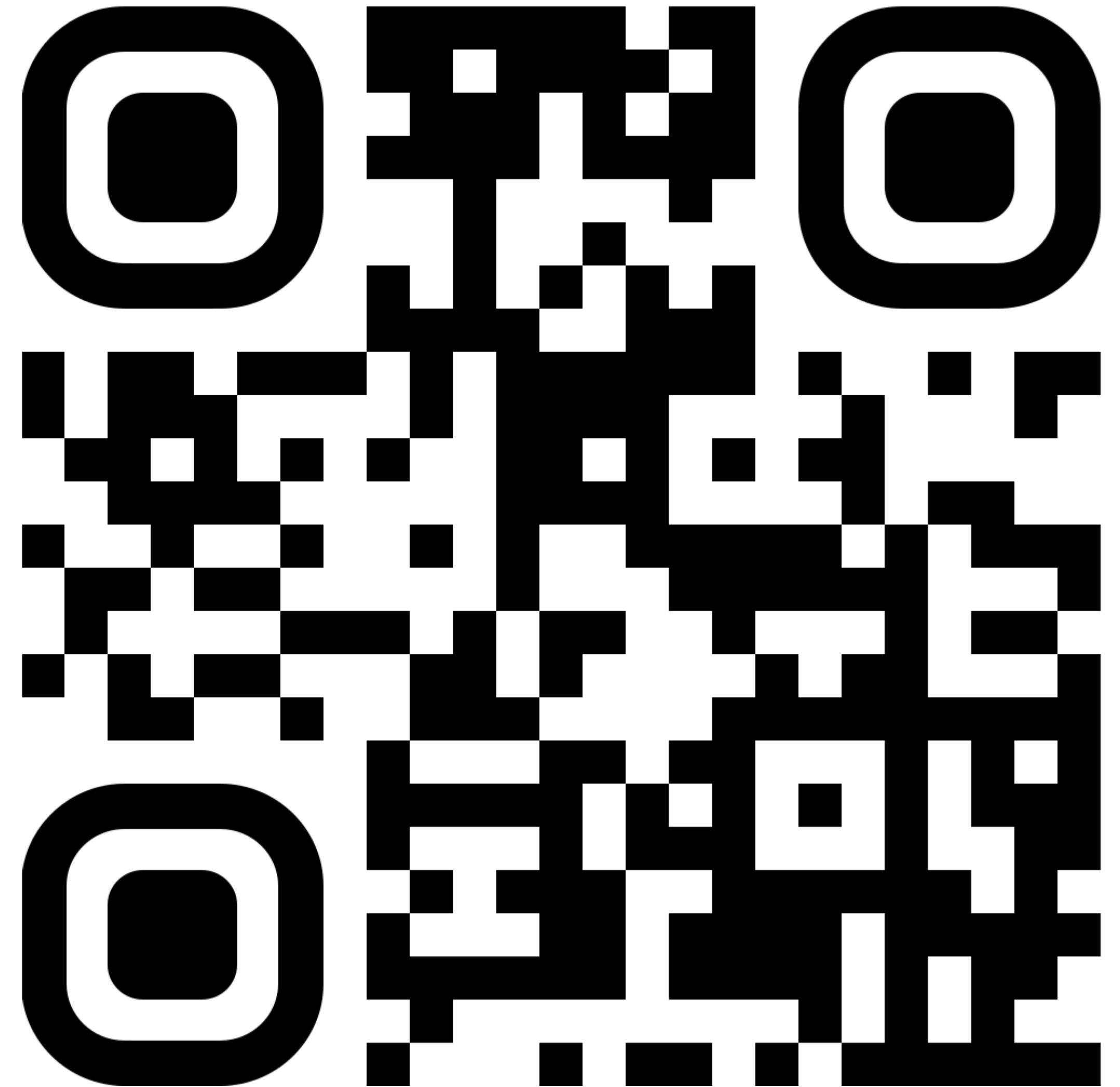
# Как работает Function Calling



# Демонстрация

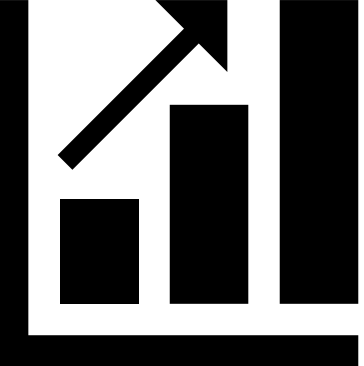
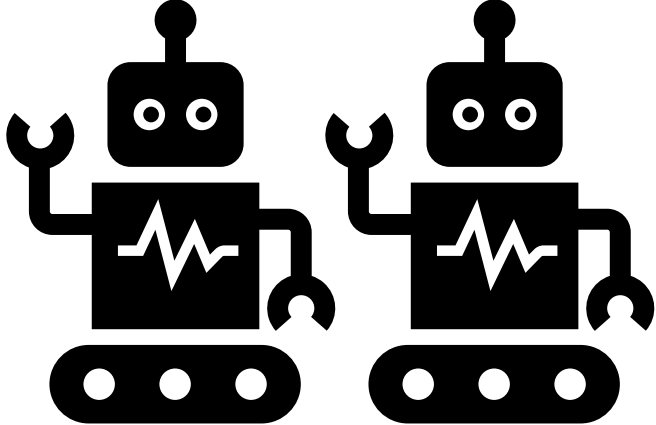
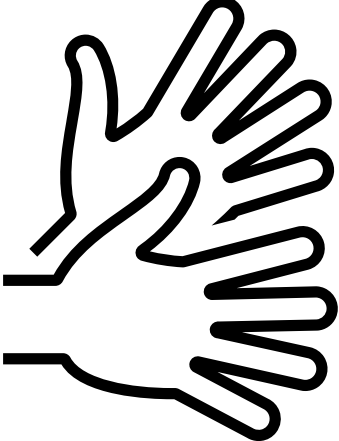


Употребление алкоголя  
вредит вашему здоровью...



[github.com/yandex-datasphere/  
advanced-assistant](https://github.com/yandex-datasphere/advanced-assistant)

# Как тестировать ассистентов?



Ручное тестирование

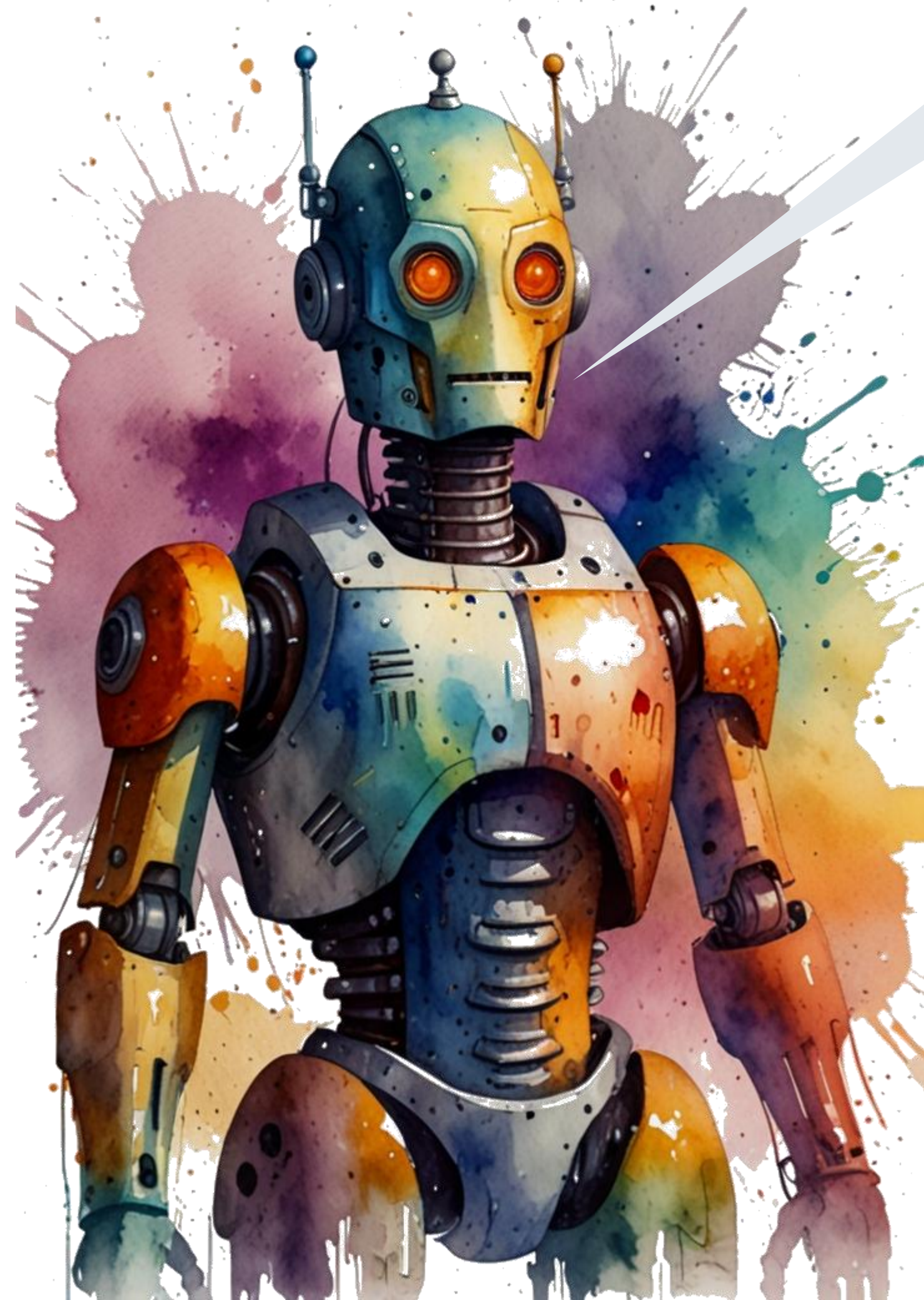
Многоагентное  
тестирование

Датасет + метрики  
RAGAS

# Многоагентное тестирование

Хочу вина!

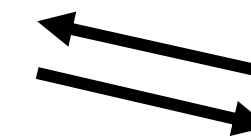
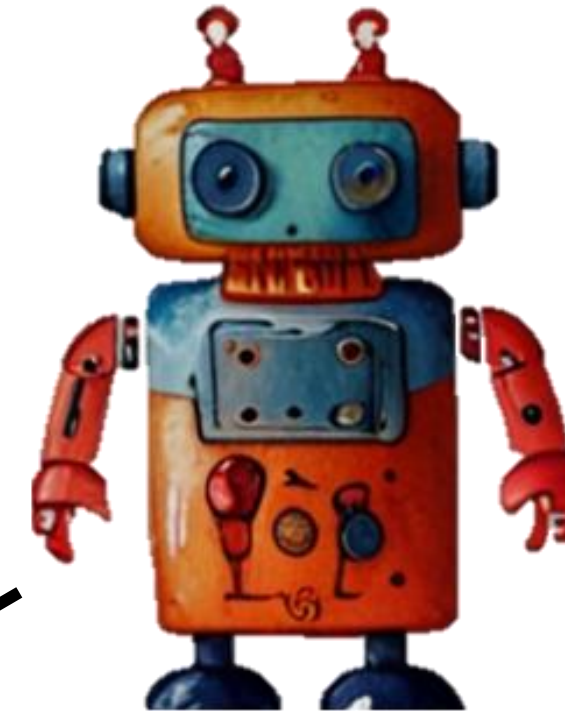
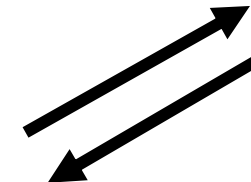
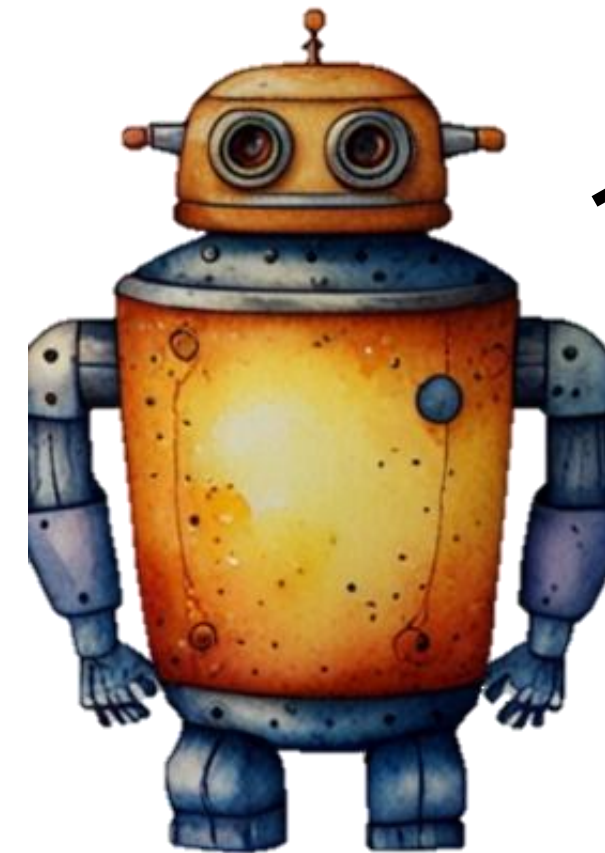
Их есть у нас!



# Выбор блюд и вина в ресторане



**Диспетчер**



**Официант:**

- \* Знает меню
- \* Передаёт запрос сомелье в случае необходимости



**Сомелье:**

- \* Знает соответствия блюд и вина

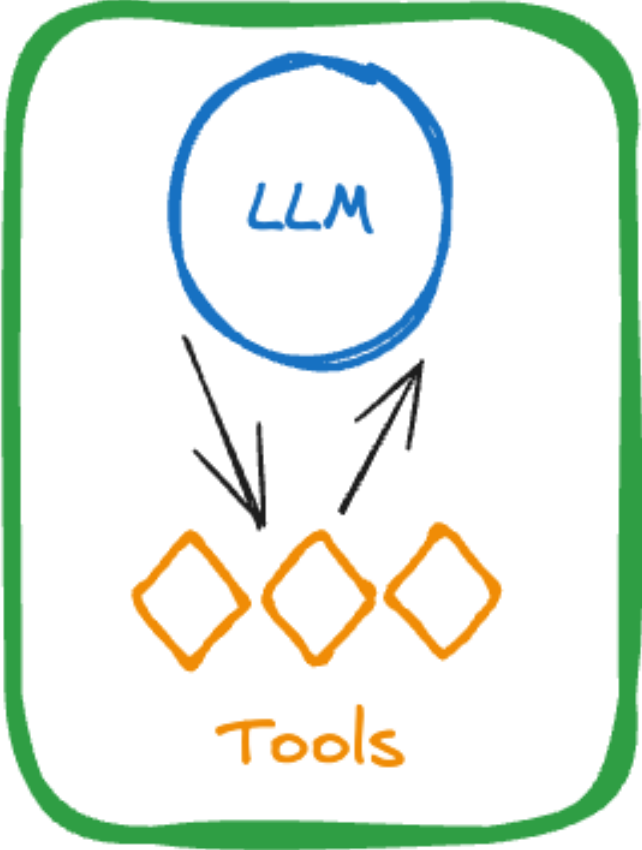
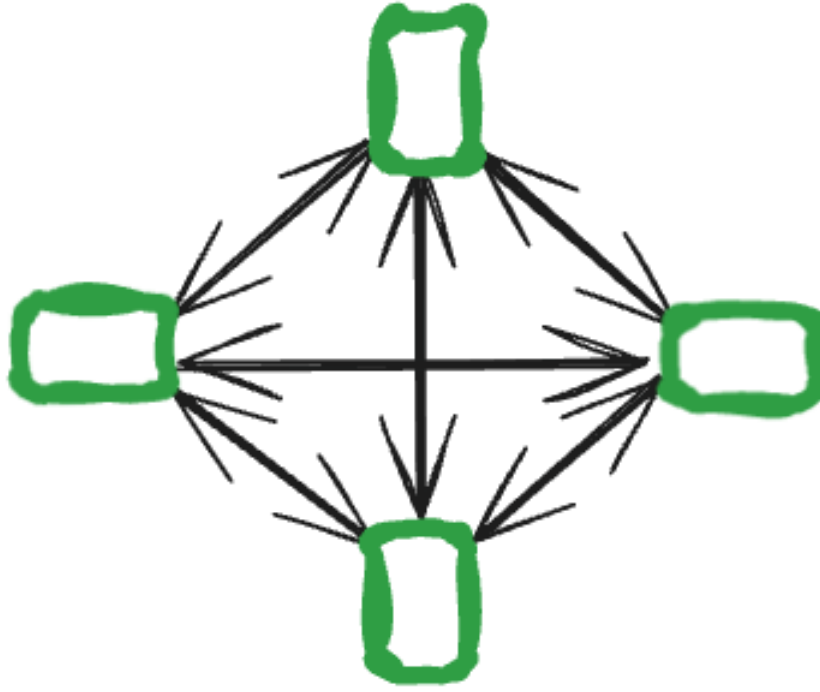
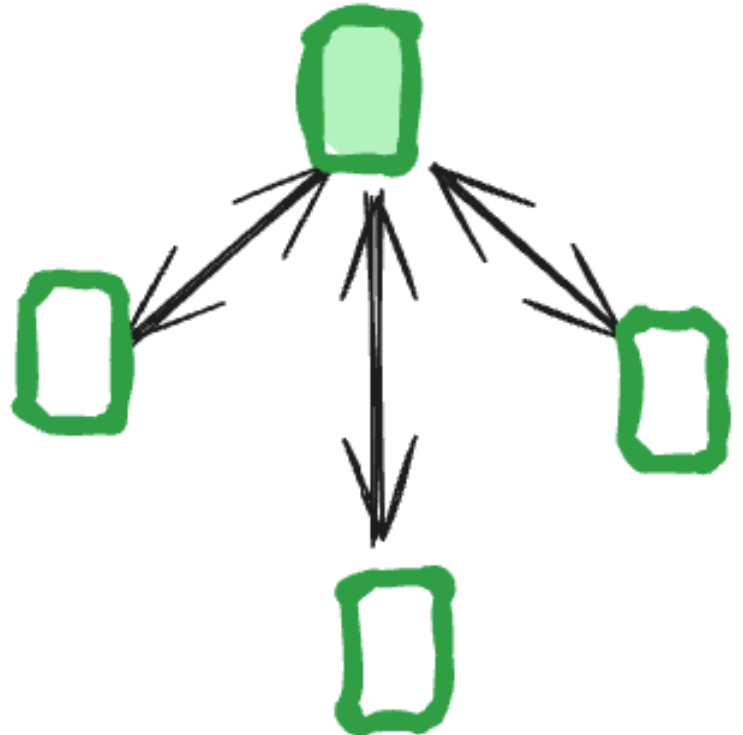
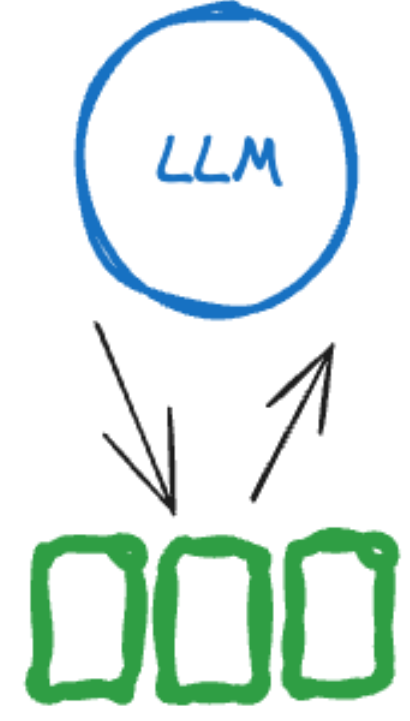
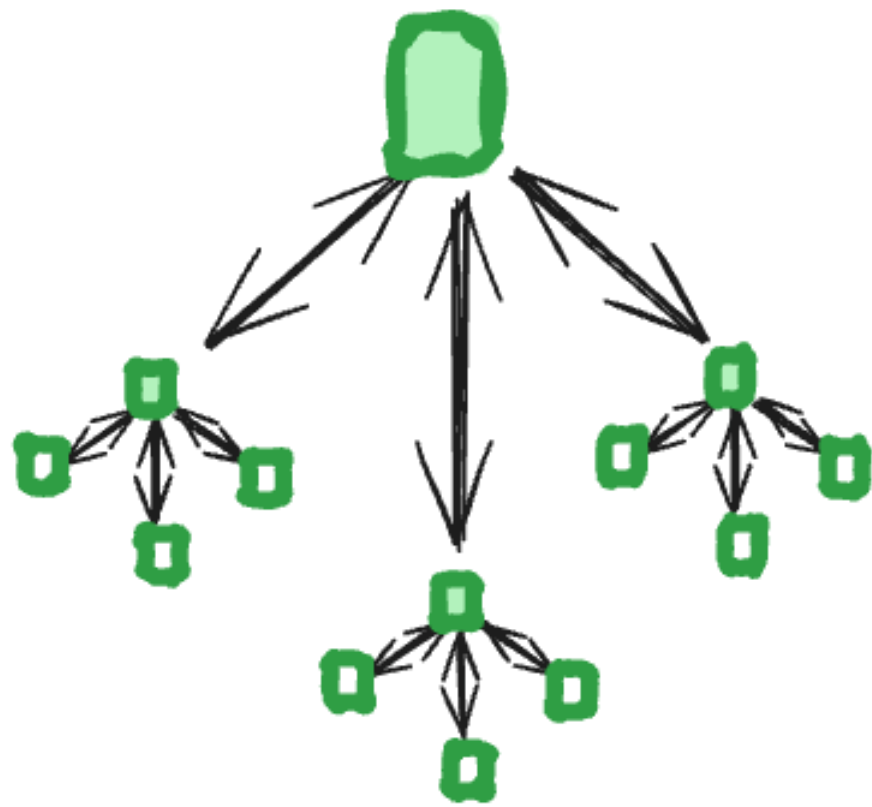
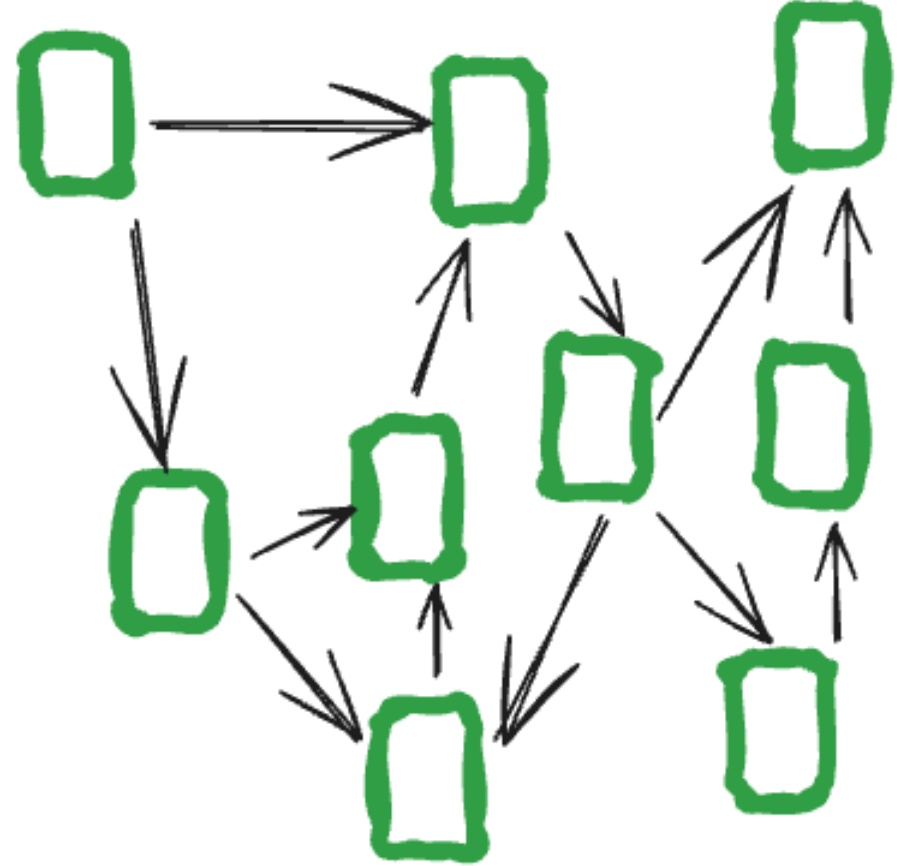


**Хостесс:**

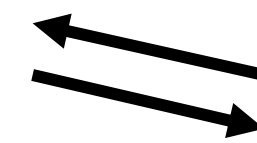
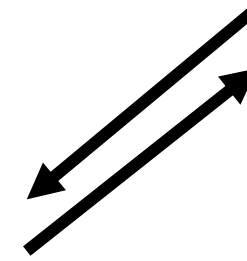
- \* Смотрит на предпочтения
- \* Передаёт дальше: официанту или сомелье



# Агентные паттерны

Single Agent	Network	Supervisor
		
Supervisor (as tools)	Hierarchical	Custom
		

# Умный (ReAct) агент



## ReAct Agent:

- Think
- Act
- Observe

## Инструменты (Tools)

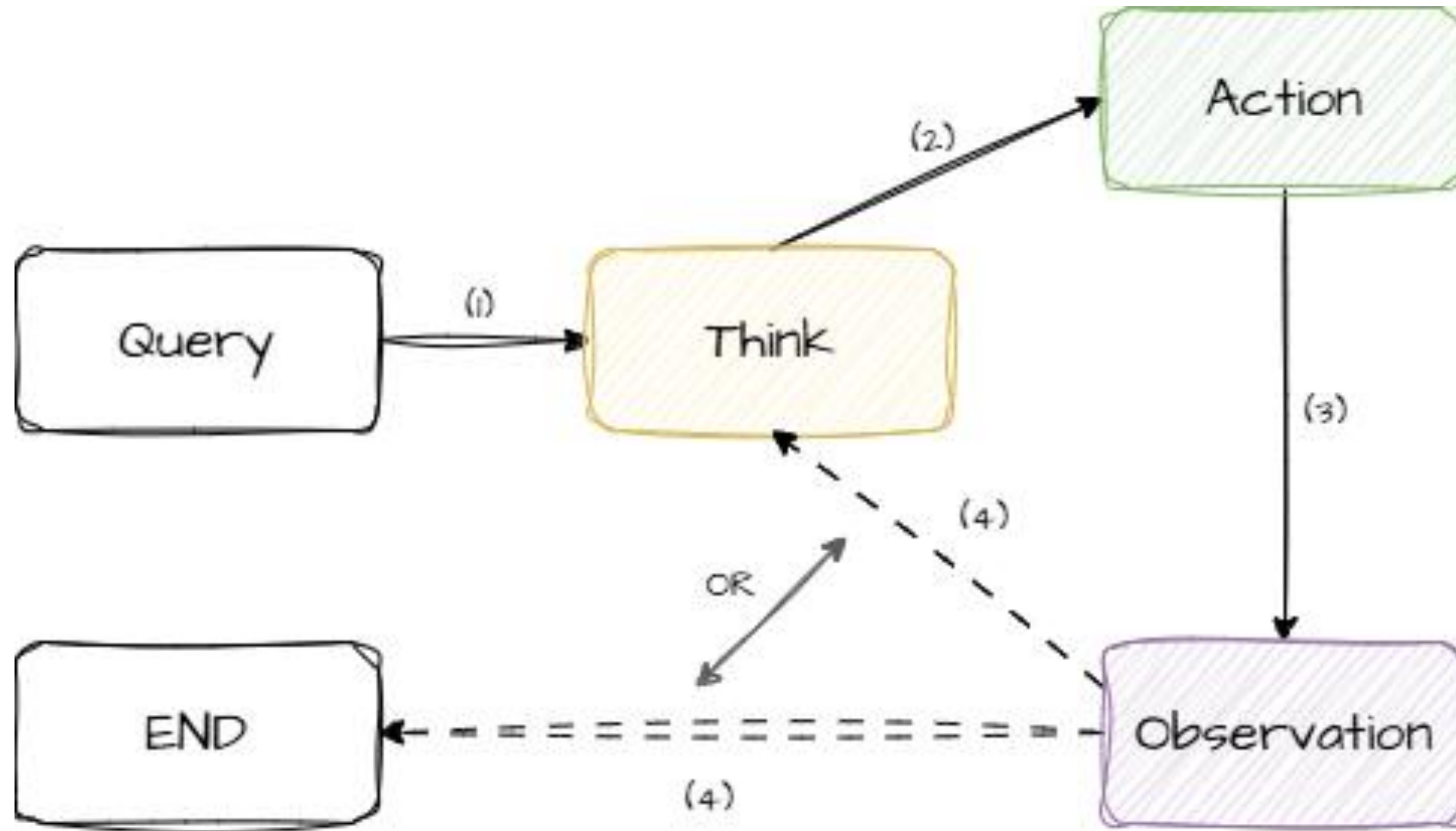
**Меню ресторана:**  
доступно удалённо по  
протоколу MCP



**Функция-  
сомелье:**  
подбор на основе  
семантического  
поиска



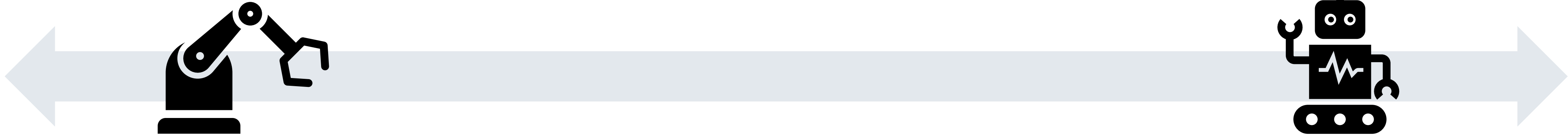
# Цикл работы ReAct-агента



# Спектр агентных технологий

Явная оркестрация

Динамическое планирование /  
кодогенерация



LangGraph  
LlamaIndex

SmolAgents

# Мораль

- 1 Агентные технологии позволяют создавать на основе LLM более сложные системы из коммуницирующих между собой автономных рассуждающих агентов
- 2 Нарботанная теория в области многоагентных систем может быть полезна: теория переговоров, аукционы.
- 3 Пока что агентные технологии очень различаются между разными фреймворками и представляют собой спектр от явной и предсказуемой окрестрации агентов, используемой в реальных индустриальных системах, до планирующих агентов с кодогенерацией, которые лучше подходят для исследовательских задач.

# Остаёмся на связи!

<http://t.me/shwarsico>



# Спасибо за внимание!



**Дмитрий Сошников**  
к.ф.-м.н., доцент МАИ и НИУ ВШЭ,  
консультант, Yandex Cloud  
[dsosnikov@yandex-team.ru](mailto:dsosnikov@yandex-team.ru)



Вступайте в сообщество  
Yandex Foundation Models  
[t.me/YFM\\_Community](https://t.me/YFM_Community)