

# Построение AI-ассистента для работы с ТИ-знаниями

Николай Арефьев



ctt

# Проблематика

- Необходимо понимать от чего защищаться, как на уровне сети, так и хоста.
- Мы сталкиваемся с различными атаками но как понять, кто за ними стоит, какова его цель и на что направлена сама атака?  
Может быть пора бежать и вырубать пограничные маршрутеры?

При каждой непонятной ситуации искать в интернете похожие инциденты?

# Источники знаний

Каждый источник привносит свою часть СТИ-знаний

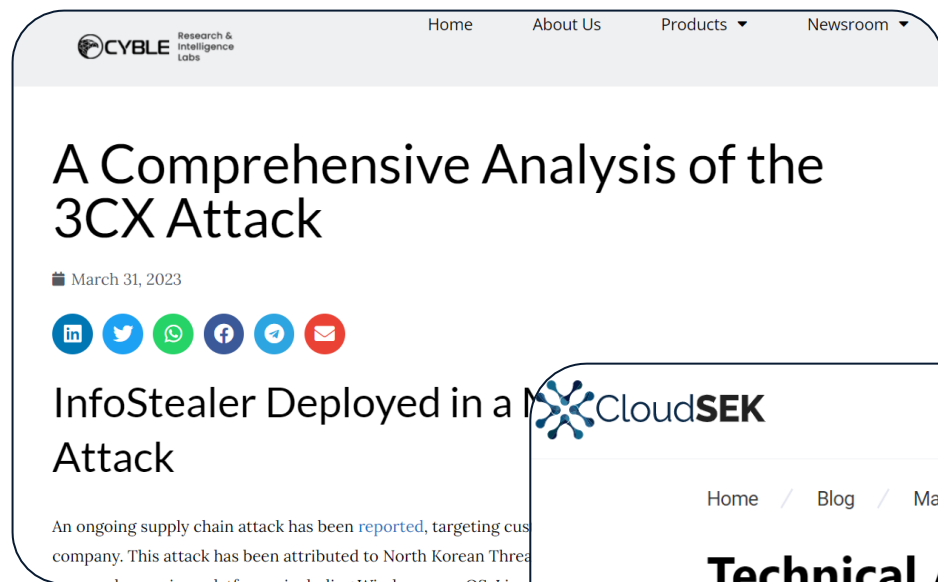


Объём данных



ctt

# Как выглядят данные

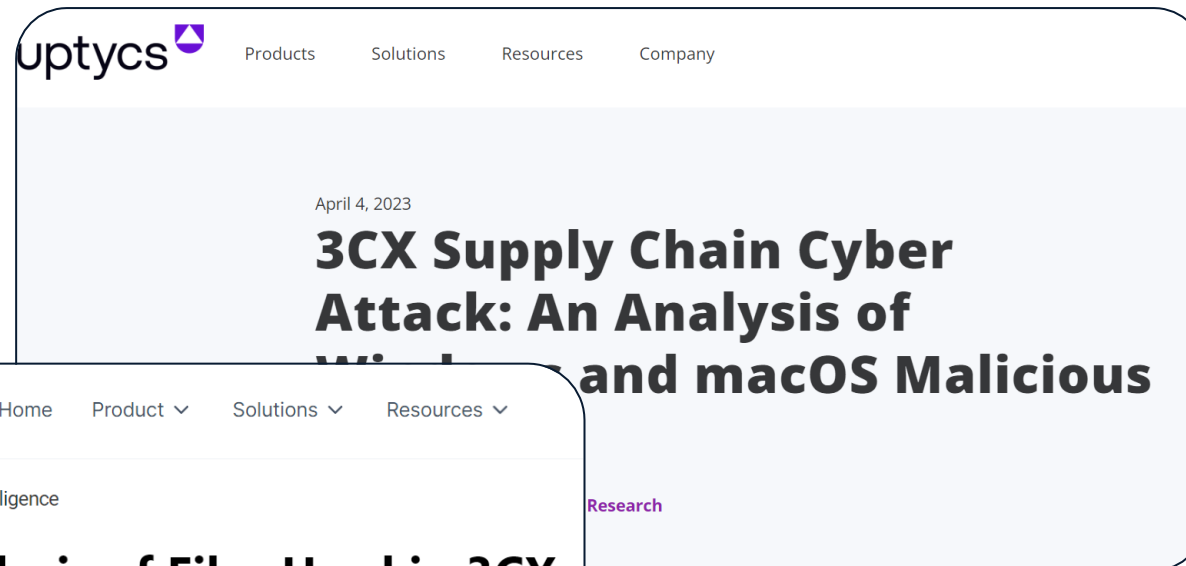


**A Comprehensive Analysis of the 3CX Attack**

March 31, 2023

InfoStealer Deployed in a Malware Attack

An ongoing supply chain attack has been [reported](#), targeting customer company. This attack has been attributed to North Korean Threat



**3CX Supply Chain Cyber Attack: An Analysis of Windows and macOS Malicious**

April 4, 2023

Research



**Technical Analysis of Files Used in 3CX Desktop App Malware Campaign**

Published: **April 4, 2023** • 8 min read

On 29th March 2023, reports emerged of malicious activities originating from a signed 3CX desktop application. Trojanized versions of the 3CX desktop app load a DLL with malicious content. The DLL launches a multi-staged attack on the victim machine, the final stage being the deployment of an unidentified info stealer.

**Mehardeep Singh Sawhney**  
[LinkedIn](#)

~4000  
ОТЧЕТОВ В ГОД

# На какие вопросы должен уметь отвечать ИБ-ассистент

- Что такое BugSleep?
- Каково поведение BugSleep?
- Как детектировать сетевую активность BugSleep?
- Были ли у нас когда-то в компании инциденты, связанные с BugSleep?

RAG

- Какие порты открыты для доступа из интернета на IP: xxx.xxx.xxx.xxx?
- Есть ли вредоносная активность на URL: <https://yyyy.yyy/zarplata.js>
- У меня есть вот такой файл, проверь, нет ли в нем ВПО?
- Есть ли сейчас на МЭ: xxx.xxx.xxx.xxx правило для разрешения доступа из подсети 10.0.0.0/24 на узел 10.120.0.1 по порту TCP:443

Multi-agents  
LLM

# На какие вопросы должен уметь отвечать ИБ-ассистент

- Если в файле ВПО, узнай что она делает, если ее активность можно заблокировать на уровне сети, **проверь** что у меня настроены все необходимые правила для **детектирования** ее сетевой активности.
- **Узнай**, сможет ли моя компания **отразить** атаку группировки АРТ УУУ.

Multi-Agents  
LLM

# Почему бы просто не спросить у ChatGPT/DeepSeek/GigaChat/YandexGPT?

- × Передача конфиденциальных данных в запросе.
- × LLM отвечает либо то, на чем была обучена, либо придумывает ответ.
- × Чем меньше обучающая выборка по теме запроса, тем больше вероятность галлюцинации.
- × Часто невозможность подкрепить ответ источником.
- × Необходимость постоянного дообучения (fine-tuning) из-за регулярно появления новых данных.

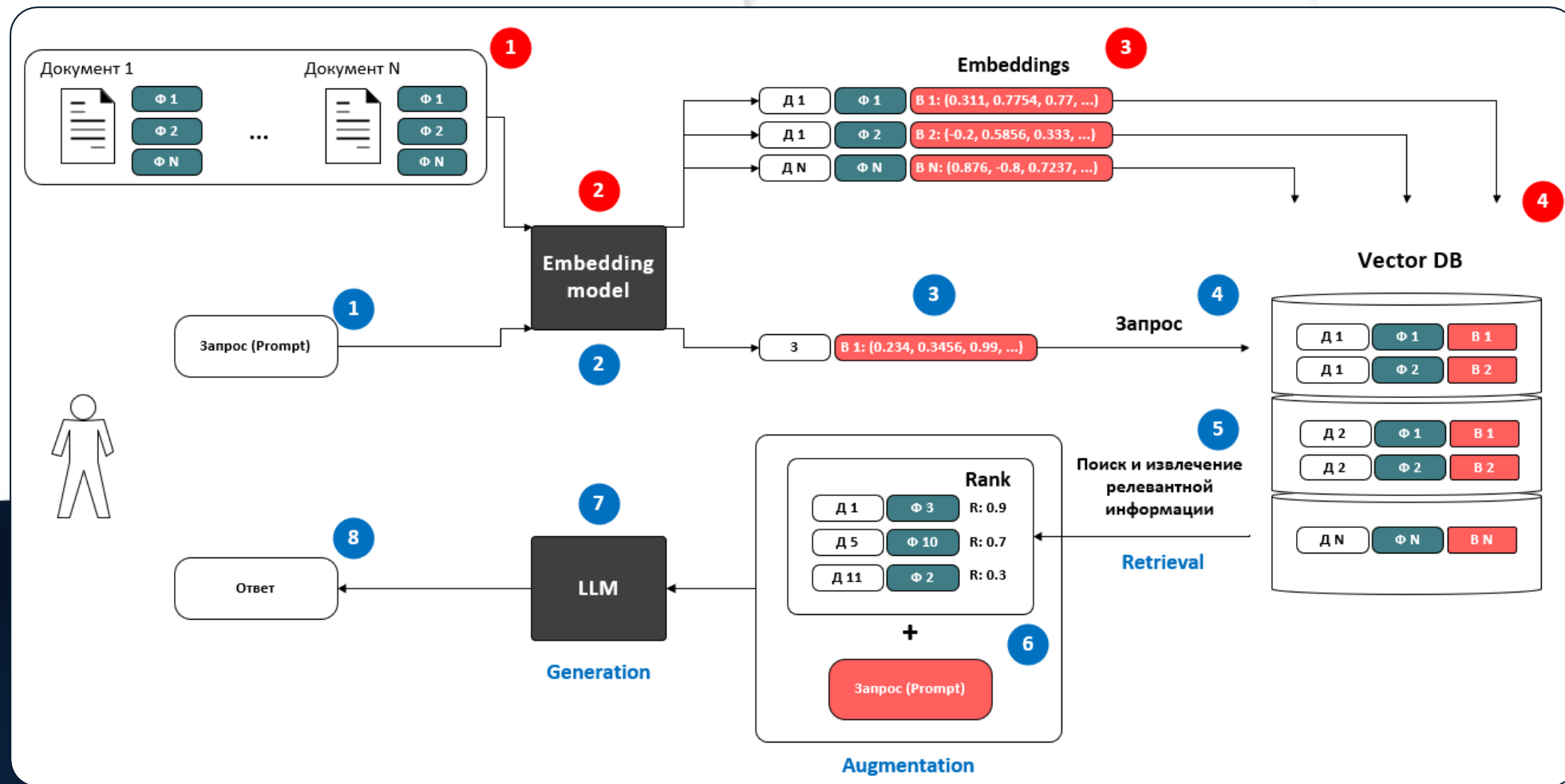
Галлюцинации

# RAG

## (Retrieval Augmented Generation)



# Общая архитектура RAG

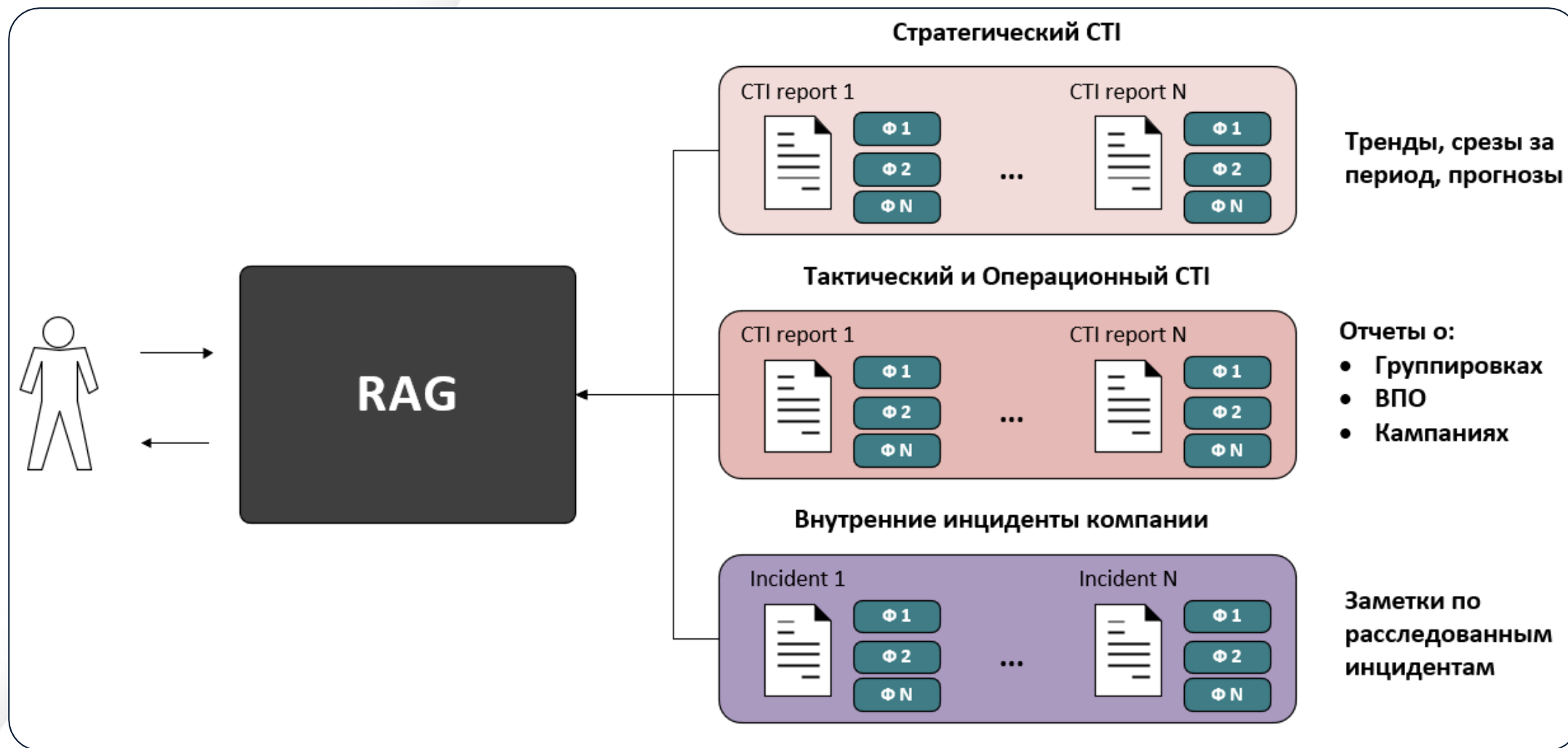


**Retrieval** - поиск и извлечение релевантной информации.

**Augmented** - дополнение найденной информацией.

**Generation** - генерация ответа с учетом найденной релевантной информации.

# Использование RAG в ИБ-ассистенте



# Примеры работы RAG

Запрос: Что такое BugSleep?

**BugSleep**, также известный под псевдонимом MuddyRot, - **это инструмент удаленного доступа** (RAT), который был идентифицирован как часть кампаний **кибершпионажа**, приписываемых **группе MuddyWater**. Впервые он был проанализирован в июне 2024...

Исходные данные:

- <https://blog.talosintelligence.com/writing-a-bugsleep-c2-server>
- <https://research.checkpoint.com/2024/new-bugsleep-backdoor-deployed-in-recent-muddywater-campaigns>

# Примеры работы RAG

## Запрос: Как детектировать сетевую активность BugSleep?

### Ключевые методы обнаружения

#### 1. Коммуникация команд и управления (C2):

- BugSleep использует специальный протокол команд и управления через простые **сокеты TCP**. Мониторинг исходящего трафика на предмет необычных подключений TCP, особенно к известным вредоносным IP-адресам или доменам, имеет решающее значение.
- Каждый маяк BugSleep **имеет определенный формат**: он **начинается с поля длиной 4 байта**, за которым следуют данные, которые обычно включают имя компьютера и имя пользователя. Это можно отслеживать на предмет шаблонов, **указывающих на активность BugSleep**.

#### 2. Правила Snort:

- Первоначальным кандидатом на обнаружение является маяк, который можно идентифицировать по его определенной структуре. Замечено, что маяк использует порт 443, что позволяет ему смешиваться с обычным трафиком HTTPS.
- Правила Snort можно улучшить с помощью flowbits для сохранения состояния в нескольких пакетах, что **позволяет более надежно обнаруживать команды**, следующие за первоначальным маяком.
- Каждая команда, отправляемая BugSleep, имеет префикс в виде 4-байтового идентификатора команды, который также можно отслеживать на предмет определенных значений, связанных с операциями BugSleep...

# На какие вопросы должен уметь отвечать ИБ-ассистент

- ✓ Что такое BugSleep?
- ✓ Каково поведение BugSleep?
- ✓ Как детектировать сетевую активность BugSleep?
- ✓ Были ли у нас когда-то в компании инциденты, связанные с BugSleep?

RAG

- Какие порты открыты для доступа из интернета на IP: xxx.xxx.xxx.xxx?
- Есть ли вредоносная активность на URL: <https://yyyy.yyy/zarplata.js>
- У меня есть вот такой файл, проверь, нет ли в нем ВПО?
- Есть ли сейчас на МЭ: xxx.xxx.xxx.xxx правило для разрешения доступа из подсети 10.0.0.0/24 на узел 10.120.0.1 по порту TCP:443

Multi-agents  
LLM

# Multi-Agents LLM (LLM-based Multi-Agents)

# Немного определений

**Мультиагентные LLM** - система, состоящая из независимых компонентов ИИ, на базе LLM, каждый из которых называется агентом, работающих вместе для достижения общей цели. Каждому агенту может быть назначена определенная задача на основе его сильных сторон \*.

\* общепринятого определение пока нет

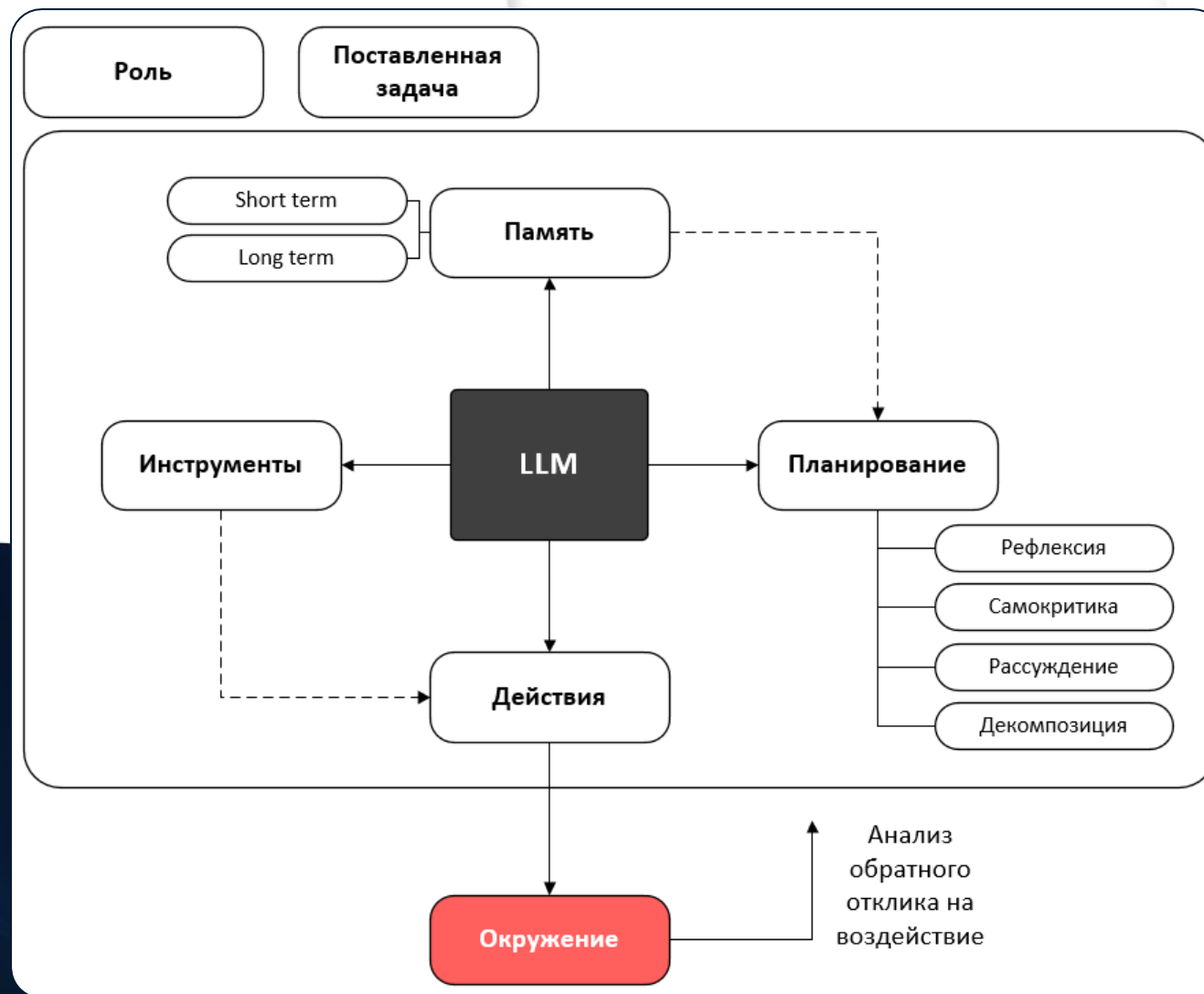
**Агент** - система ИИ, которая использует LLM для взаимодействия с окружением [1].

**Агент** - программа, которая использует методы ИИ для выполнения задач, которые обычно требуют человеческого интеллекта [2].

Источники:

1. Cognitive Architectures for Language Agents (<https://arxiv.org/pdf/2309.02427>)
2. TPTU: Large Language Model-based AI Agents for Task Planning and Tool Usage (<https://arxiv.org/pdf/2308.03427>)
3. Are multi-LLM-agent systems a thing? Yes they are. But. (<https://gist.github.com/yoavg/9142e5d974ab916462e8ec080407365b>)

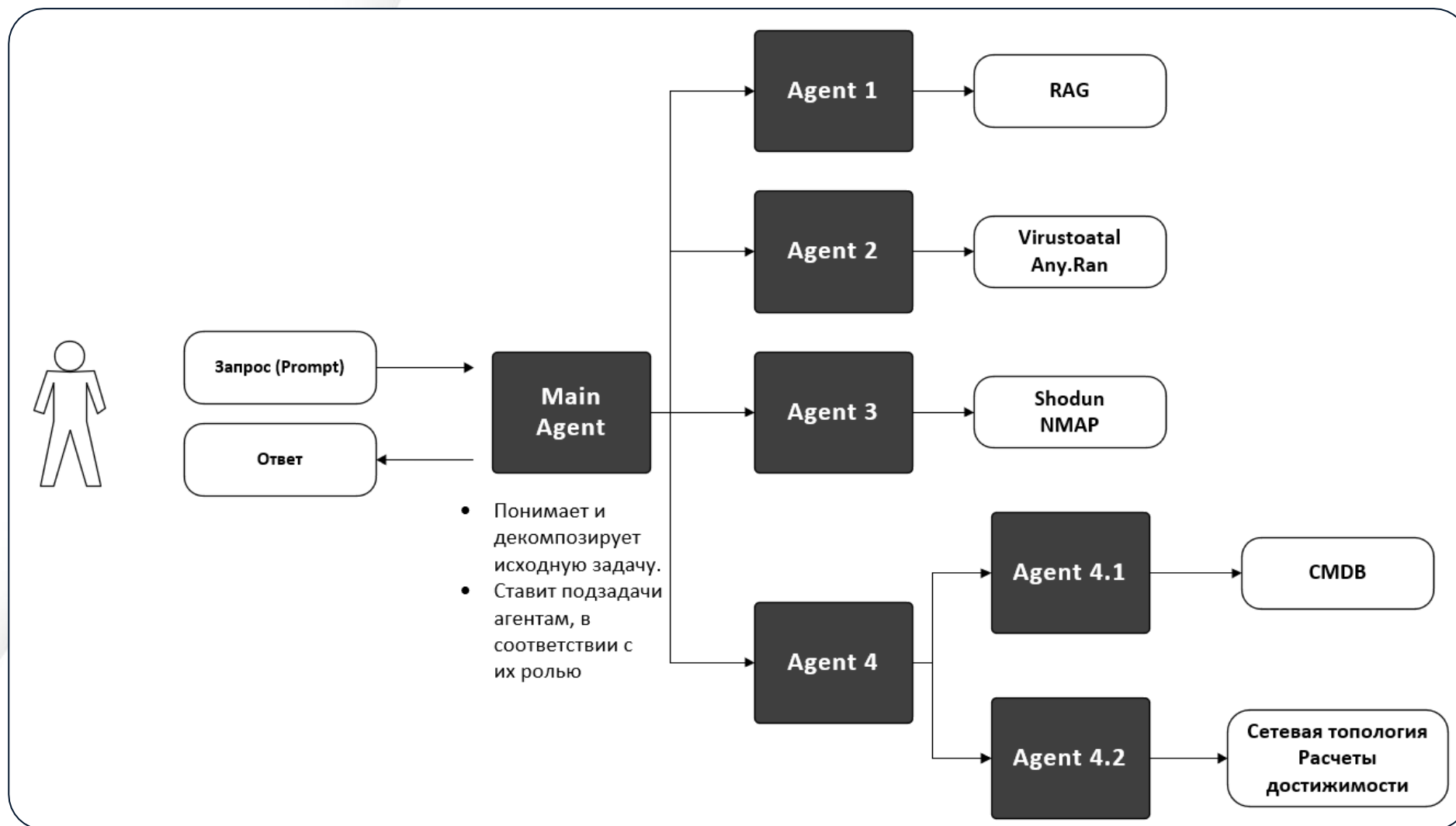
# Общая архитектура Агента



TPTU: Large Language  
Model-based AI  
Agents for Task  
Planning and Tool  
Usage



# Использование Multi-Agents LLM в ИБ-ассистенте





# Спасибо за внимание



Оставить  
фидбек по  
докладу



# ctte