

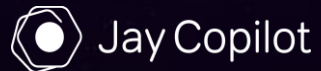
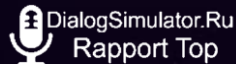
Надежные агенты

target ai



AI: помогать или заменять?

Сделать оператора
эффективнее



Автоматизировать работу
оператора



Выше эффект

Выше ответственность



Операторы не только отвечают на вопросы



30%

обращений можно решить
ответами по базе знаний

70%

требуют действий с
учетными системами

Сделать сложные интеграционные сценарии с помощью сценария очень трудоемко

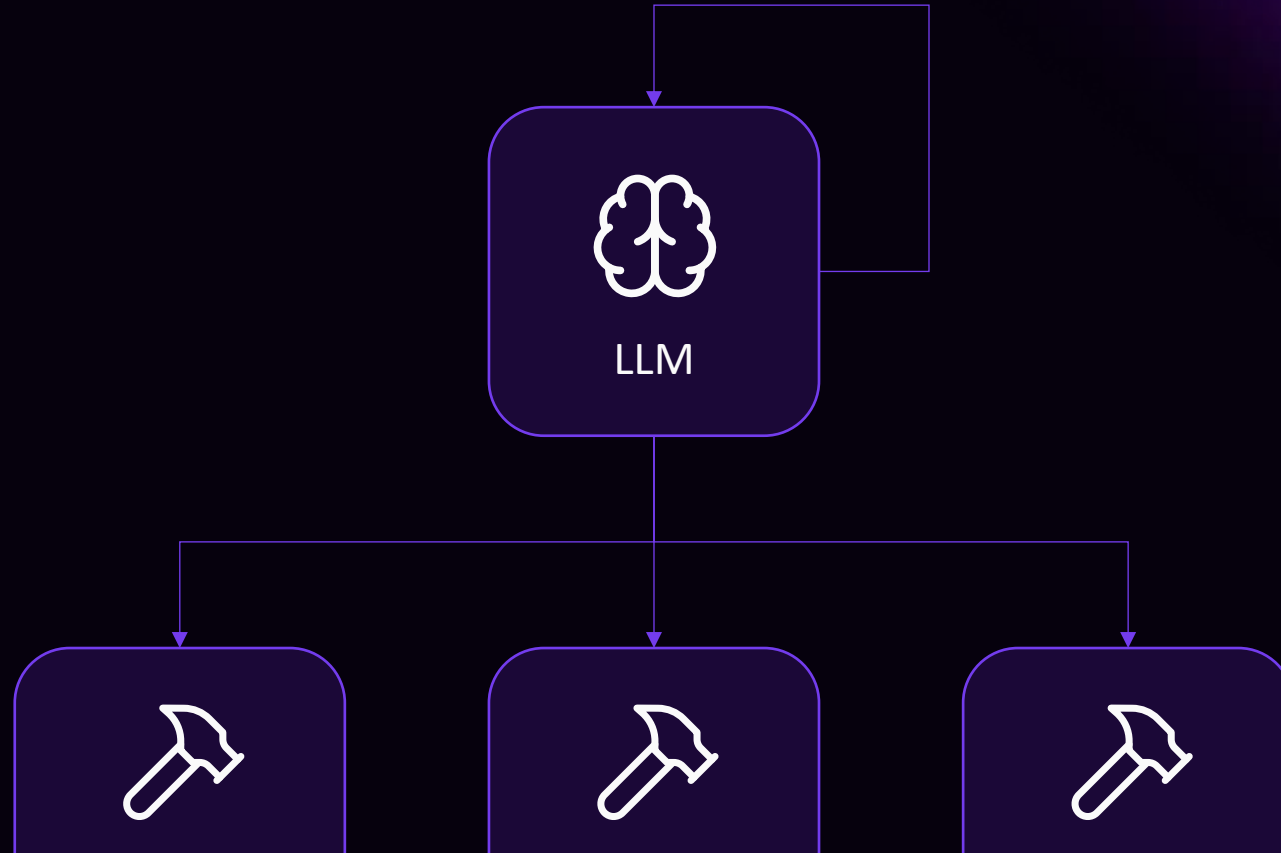


Упростим задачу с LLM





Упростим задачу с LLM



Агент – это зацикленная LLM с подключенными инструментами (интеграциями) *



Агент ≠ хороший промт

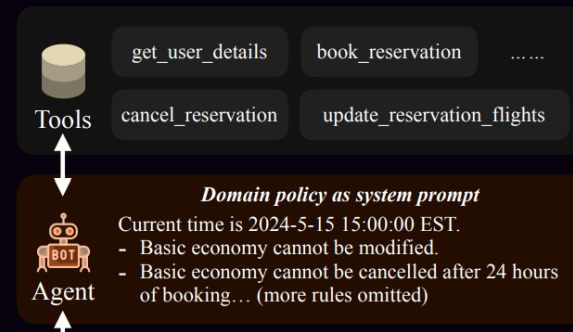
GPT-4o справится с возвратом товара в 60% случаев, даже если дать ей все инструкции*

τ -bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains

Shunyu Yao* Noah Shinn Pedram Razavi Karthik Narasimhan

Sierra

(a) τ -bench setup



Model	retail	airline	avg
<code>gpt-4o</code>	61.2	35.2	48.2
<code>gpt-4-turbo</code>	57.7	32.4	45.1
<code>gpt-4-32k</code>	56.5	33.0	44.8
<code>gpt-3.5-turbo</code>	20.0	10.8	15.4
<code>claude-3-opus</code>	44.2	34.7	39.5
<code>claude-3-sonnet</code>	26.3	27.6	27.0
<code>claude-3-haiku</code>	19.0	14.4	16.7
<code>gemini-1.5-pro</code>	21.7	14.0	17.9
<code>gemini-1.5-flash</code>	17.4	26.0	21.7
<code>mistral-large</code>	30.7	22.4	26.6
<code>mixtral-8x22b</code>	17.7	31.6	24.7
<code>meta-llama-3-70B</code>	14.8	14.4	14.6

* Yao, S., Shinn, N., Razavi, P., & Narasimhan, K. (2024). τ -bench: A benchmark for tool-agent-user interaction in real-world domains. arXiv. <https://arxiv.org/abs/2406.12045>



Проблема: LLM ненадежны

В бизнес-процессе 10 шагов

Модель успешно проходит шаг в 95% случаев

Бизнес-процесс будет полностью выполнен в $0.95^{10} \approx 60\%$ случаев

Как достичь 100% выполнения каждого шага?



Решение

Экзоскелет для LLM



Скелет, который поддерживает LLM для выполнения строгой бизнес-логики мы называем **процедурой**



Как работает процедура



«Мне нужно отменить заказ»



Спроси, почему клиент хочет отменить

«Подскажите пожалуйста, почему вы хотите отменить заказ?»



«Да товара нужного нет смысл тогда»



Уточни, какого именно товара нет

«Подскажите, а какого именно нужного товара не оказалось?»





Как работает процедура

Процедура динамически управляет диалогом: выбирает нужный блок бизнес-процесса и инжектирует команду из него в историю диалога.

«Нужно отменить заказ»

Спроси, почему клиент хочет отменить

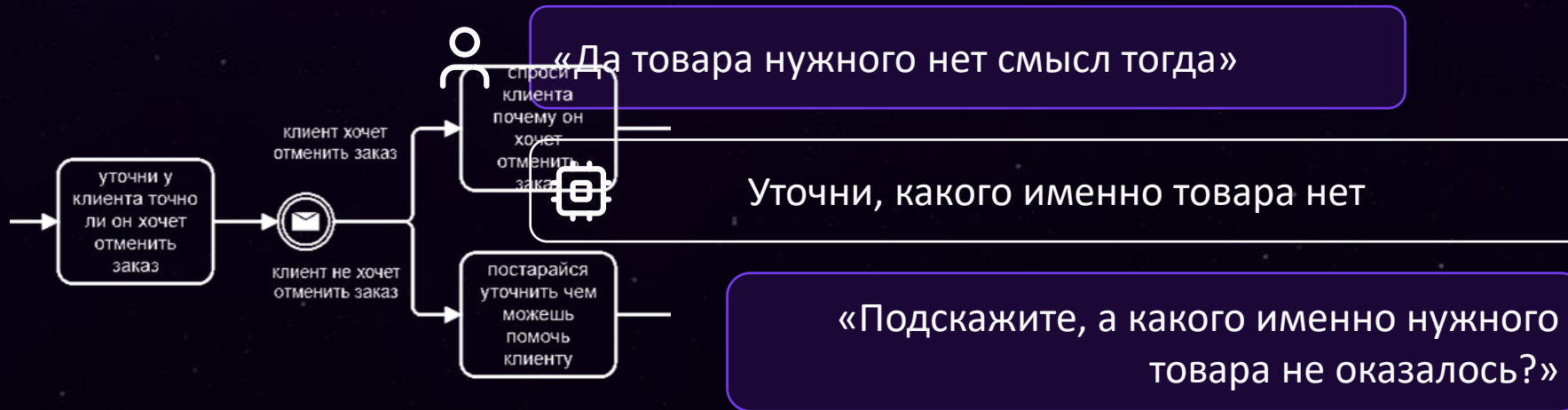
«Подскажите пожалуйста, почему вы хотите отменить заказ?»



«Да товара нужного нет смысл тогда»

Уточни, какого именно товара нет

«Подскажите, а какого именно нужного товара не оказалось?»





Эволюция агента





Эволюция агента





Процедуры делают агентов надежнее

Строгое соблюдение
бизнес-процесса

Встречные вопросы
для RAG

Выполнение регламентов

Исключение галлюцинаций

Принудительный вызов
функций



Бизнес-нотация без
ключевых слов, NLU

Поддерживают текстовые,
неточные условия

Могут прерываться и
продолжаться

Вызываются агентом по
собственному желанию

Допускают размышление
агента «про себя»

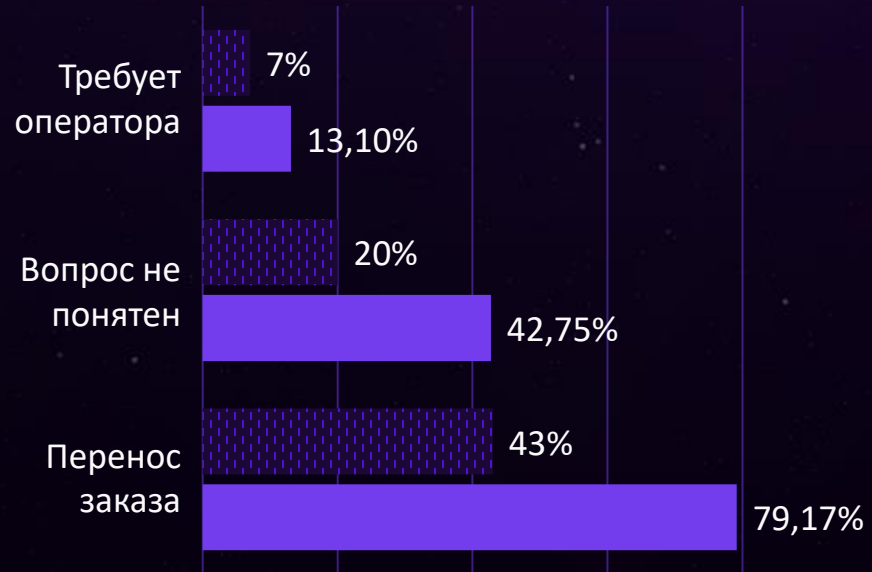
**При этом
сохраняют
гибкость**



Проверено на 87 000 диалогов в 3 сценариях

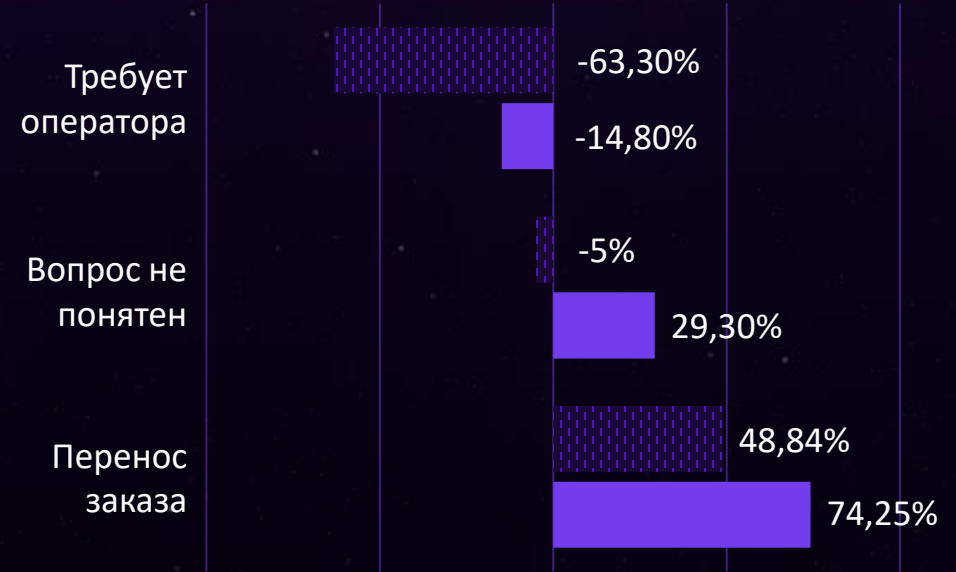
Робот Агент

Автоматизация



1.95x влияние на автоматизацию от внедрения агента

CSI



+36 п.п в среднем рост CSI от уровня робота



Технологии

Речевые технологии

Синтез и распознавание речи,
Brand Voice от **Yandex @ Cloud**

Open-source LLM

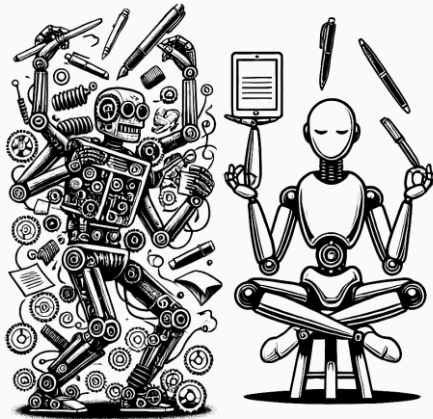
Большая языковая модель

Инфраструктура

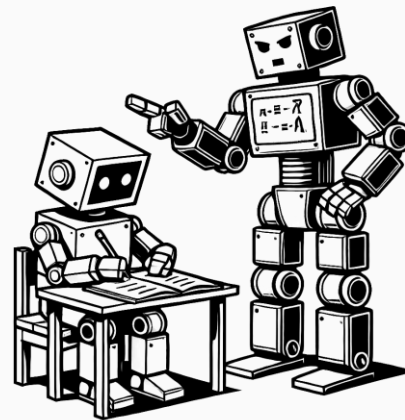
Виртуальные машины в
Yandex @ Cloud

Агенты уже заменяют роботов

LLM-агенты быстрее и проще роботов решают транзакционные задачи



Процедуры заставляют агента строго выполнять бизнес-процесс



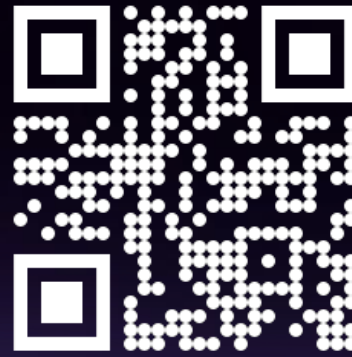
Внедрение агентов кратно повышает автоматизацию и CSI



Сделаем умных, надежных агентов в **target ai**, приходите!



targetai.ai



[@glebdementev](https://twitter.com/glebdementev)