Команда Сбера | Robotics Center

Physical AI

Что отделяет нас от роботов общего назначения

Святослав Морошкин · Сбер · Joker 2026

Санкт-Петербург · 2026

Святослав Морошкин

15 (21) лет в робототехнике

Управляющий директор
Центра робототехники Сбер

Делаем гуманоидного робота Грин
и разрабатываем Physical AI



Центр Робототехники

Профессионалы всех направлений PhysicalAI

- Роботы общего назначения
- Роботизация логистики
- Универсальные Action-модели



2017 — Создание Лаборатории робототехники Сбера

2019 — Лаборатория → Центр

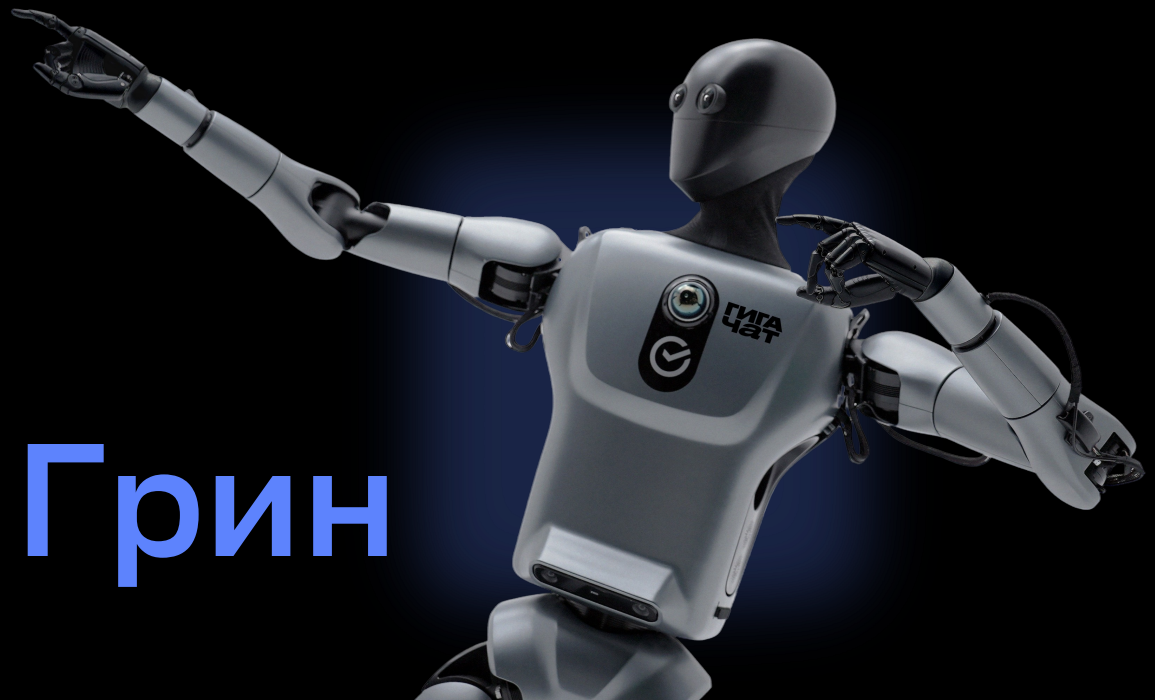
2023 — Решение о разработке антропоморфных роботов

2025 — Презентация Грина на AIJ'25

Сейчас — Развитие PhysicalAI

Наши роботы

Флагманский антропоморфный робот.
Автономно ориентируется в пространстве,
выполняет простые задачи, говорит
на естественном языке.



Грин



Полуантропоморф
для ритейла,
складов, автопрома

Воркер



Робособака
для городской
доставки

Курьер

Что вы унесёте за 45 минут

ИНЖЕНЕРУ

Робот набор систем – от хардвера до ИИ, каждая вносит вклад в способности

В целом, каждая по-отдельности вполне себе осваиваема

Рзаберемся. где фронт работ, открытые задачи

ПРОДАКТУ

Четыре среза, которые отделяют demo от production:

1. OOD success rate — что в видео vs что на новых сценах
2. Длина задачи vs reliability — на скольких шагах разваливается
3. Data hunger — сотни или тысячи демо на новый навык

Почему "купите домой через год" — маркетинг, не roadmap

ЗАИНТЕРЕСОВАННОМУ

Ответ на «где мы на самом деле находимся»

И *неочевидный* прогноз, к которому вернемся в финале



Купили бы вы себе такого робота домой?


3X

А такого?

Не надо думать про робота, как про целое

Быстренько разберемся со стеком технологий, тут немного



AI

Pretrain Data
- OpenSource Robots Data
- Teleop Data
- Human Data
- Synthetic Data

Models
- Perception
- VLM/ Reasoning
- VLA
- RL / IL WBC Control Policy
- World Model

Physical AI Factory
- Learning Infra (GPUs + Network)
- Data Collection & Management System
- ML Training Pipelines
- General Sim

FineTune Data
- Teleop Data
- RL Finetune

Software

Robot Software
- Robot Platform API
- OS
- Sensor Data Streaming
- Middleware
- Utilities
- Low level Safety
- State Estimator
- Diagnostics & Logging
- Configurator

Software Infra
- Software CI/CD Pipelines
- Cloud platform
- CyberSecurity
- Fleet Management System

Hardware

Robot Hardware
- Links
- Exterior
- Power Mgmt
- Motor Drivers
- Communication (Can2Eth)
- Motors
- Dexterous Hands
- Computers
- Cameras
- Encoders
- IMU
- Tactility

HW Infra
- Robot Testing Facilities
- Опытное производство
- Supply Chain
- Reliable Robots Fleet
- Support Team
- HIL + SIL Stands
- Whole Body Teleop Device/ Technology

Робот = 4 слоя

LAYER 4 — **HIGH LEVEL REASONING: VLM**

LAYER 3 — **MOTION PLANNING: VLA**

LAYER 2 — **MOTION CONTROL: RL WBC**

LAYER 1 — **HARDWARE: Приводы, сенсоры, батареи, HAL**

Надо чем-то двигать: Hardware

В основе всего - электропривод



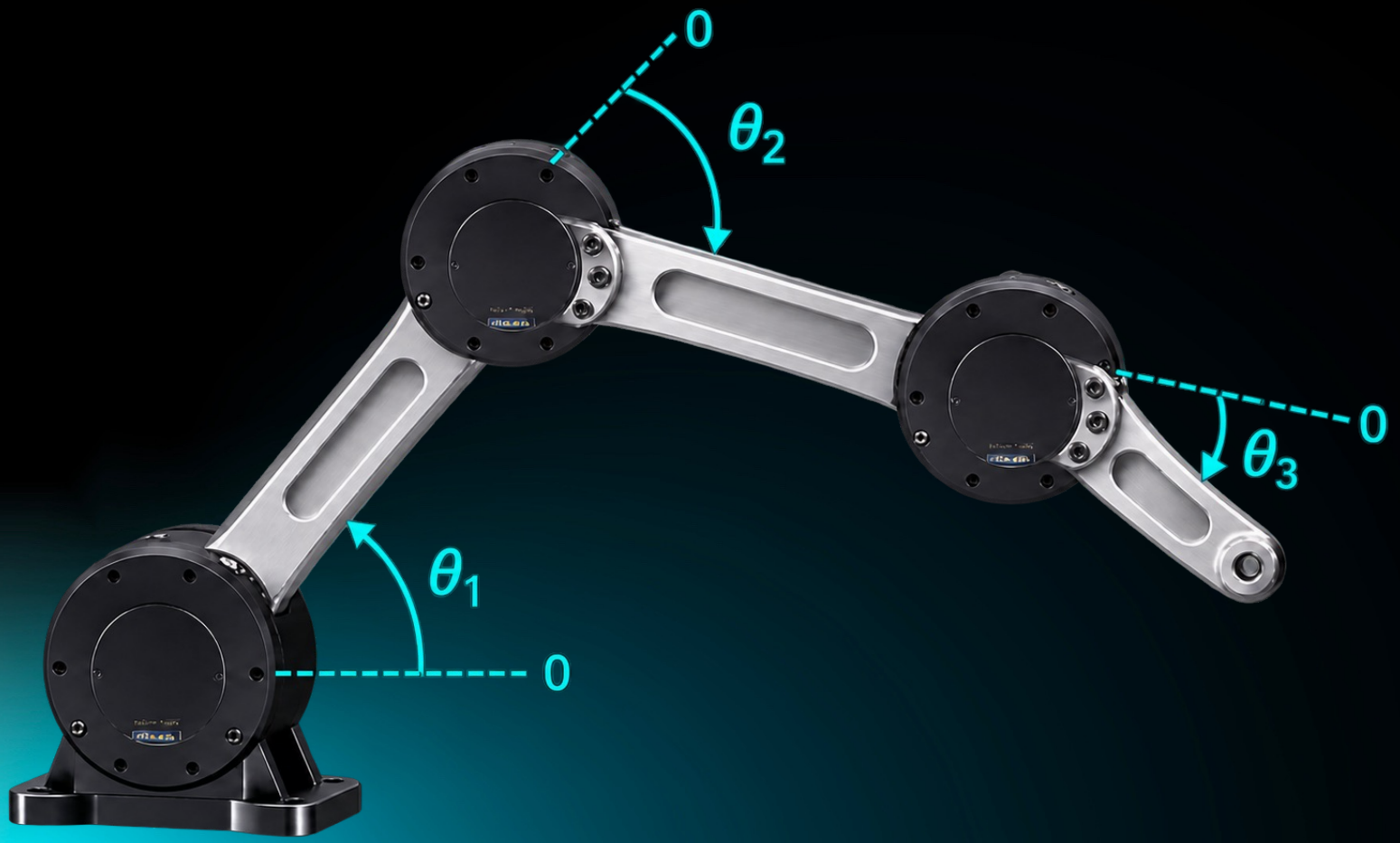
Драйвер

Электро
мотор

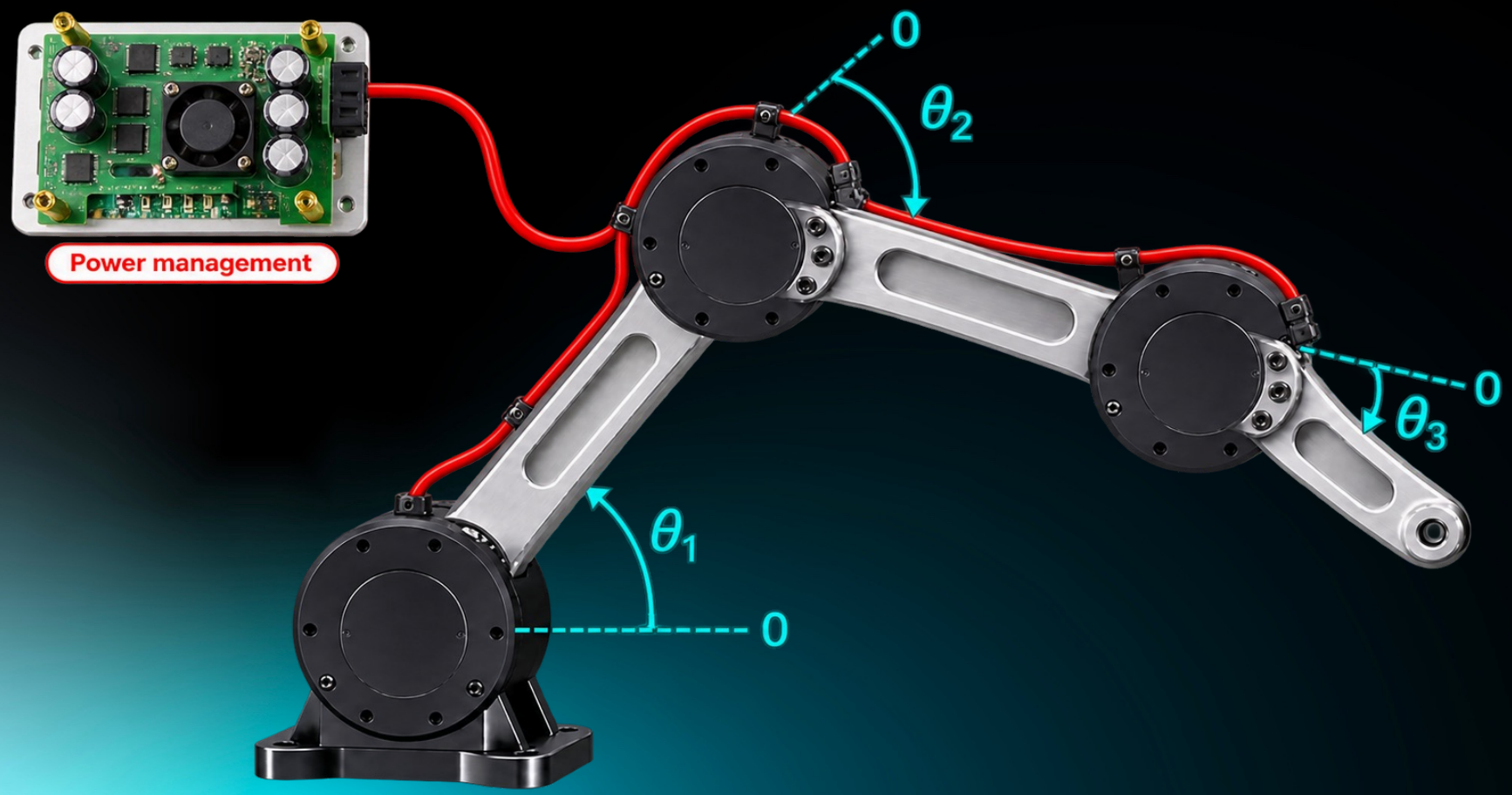
Редуктор

Корпус

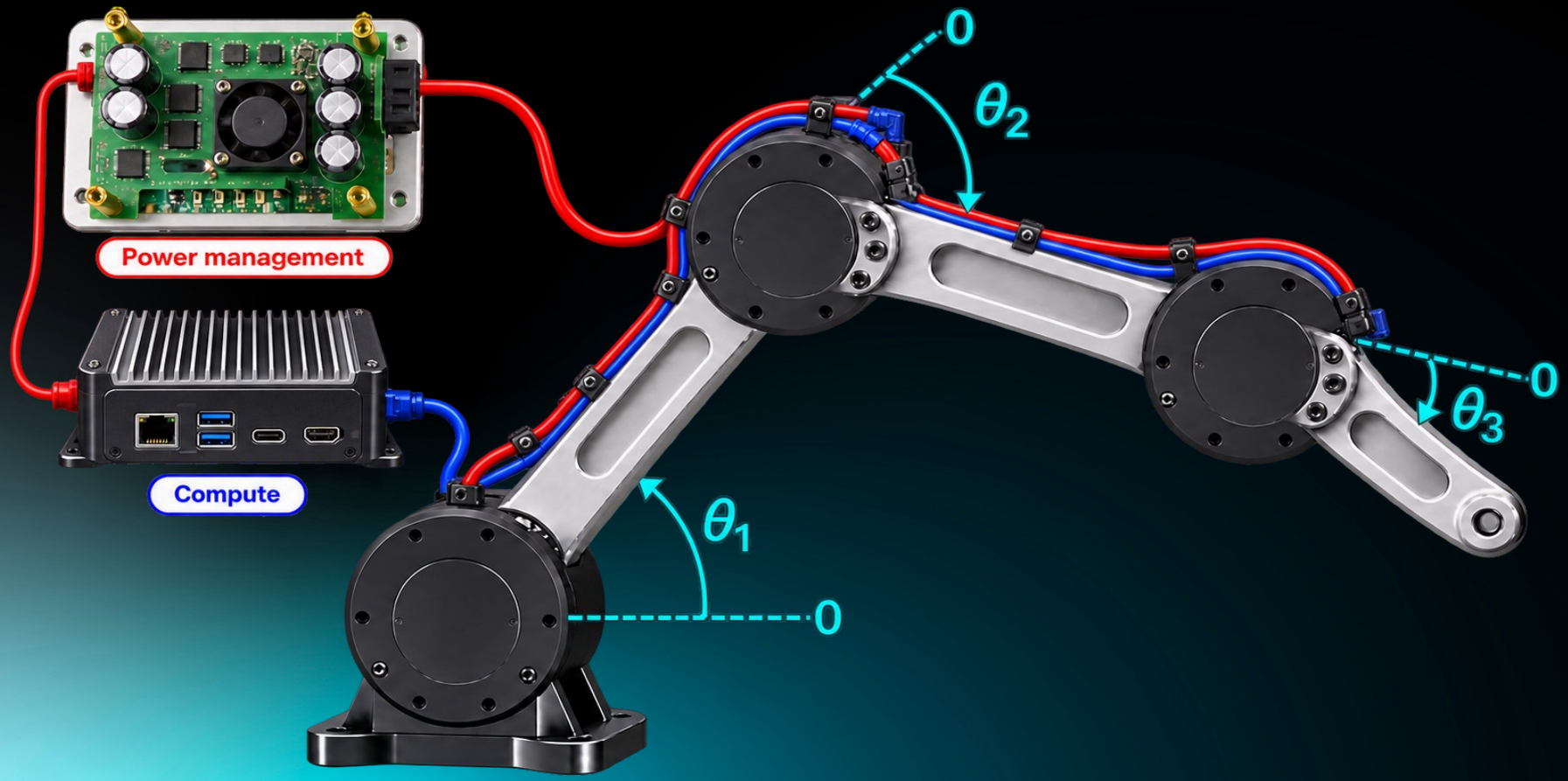
Соединяем несколько моторов – получаем простейший манипулятор



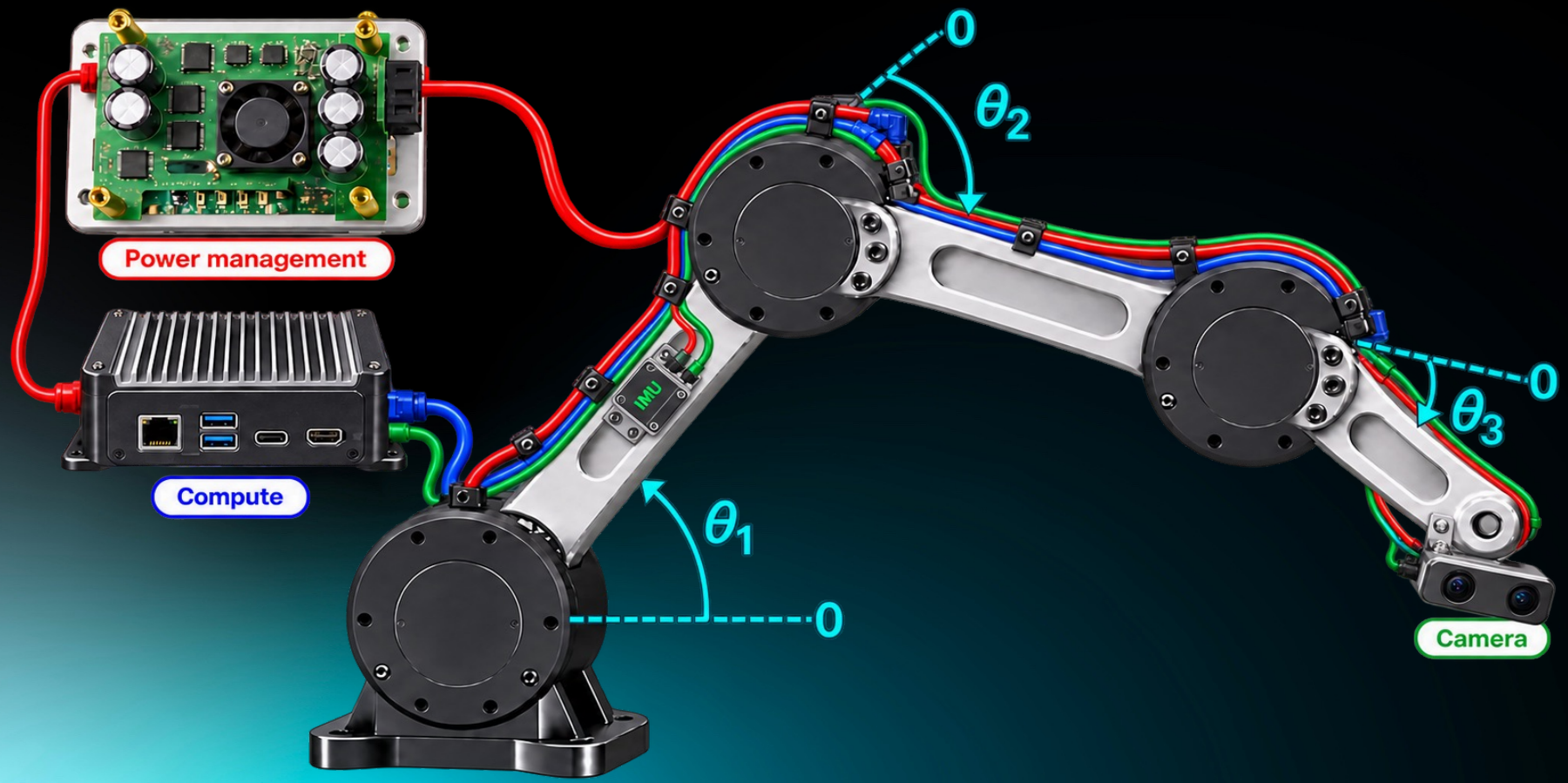
Добавляем питание



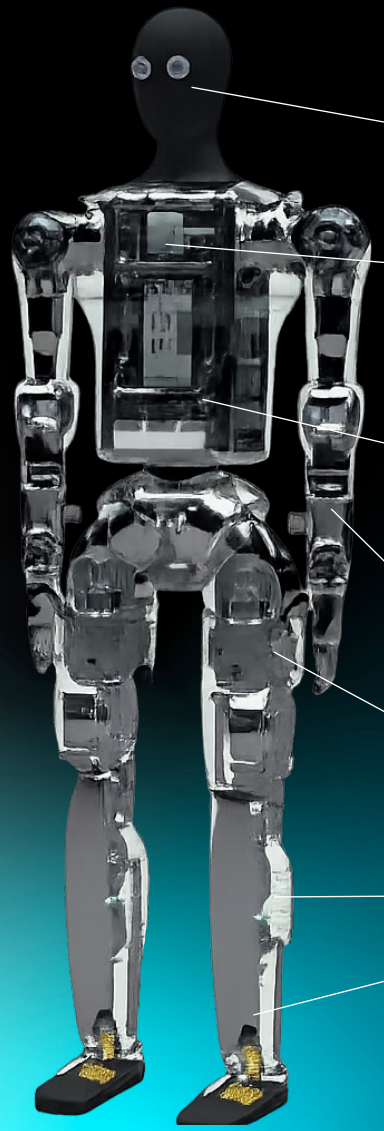
Добавляем управление



Добавляем сенсорику



Соединяем все это скелетом





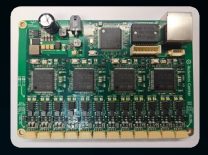
Lidars



Cameras



Power boards



Onboard routers



Control Boards



Onboard compute

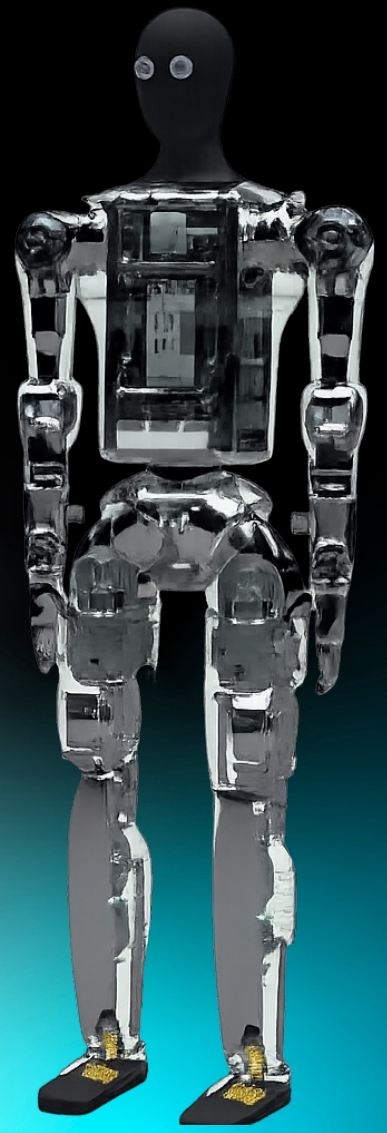


Motors



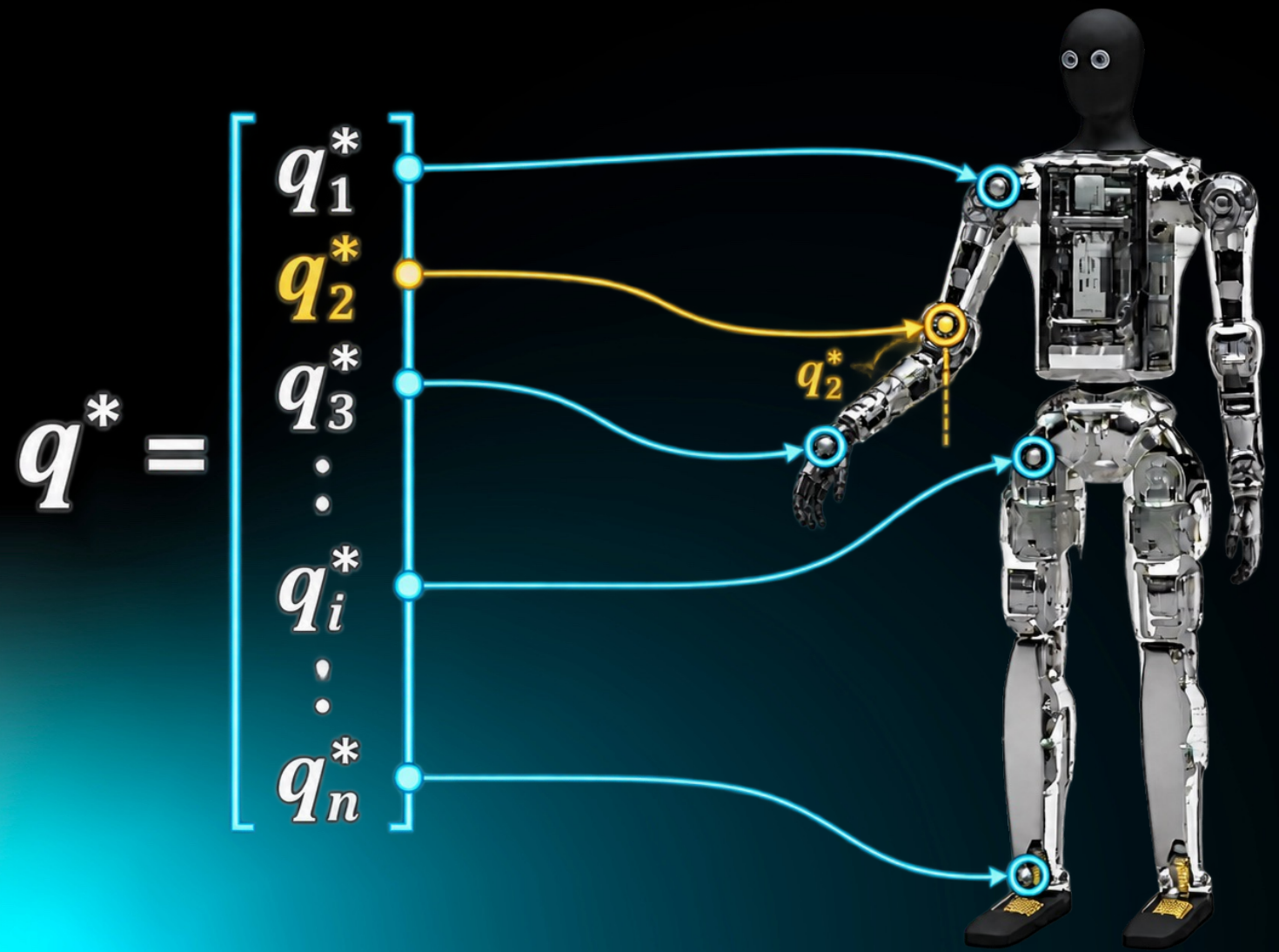
… и ~~не~~много софта

Интерфейс наружу: «согни локоть на N градусов»

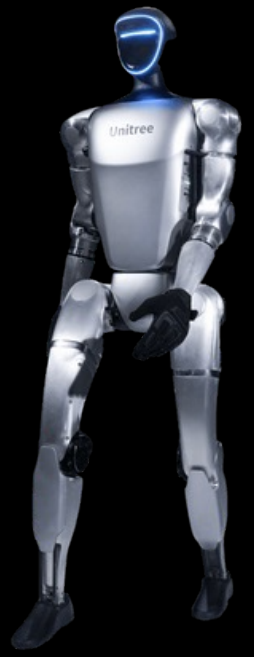


+ hardware abstraction layer
+ system services
+ drivers

Получился робот с интерфейсом «согни локоть на N градусов»



За последние 3 года произошел бум аппаратных платформ



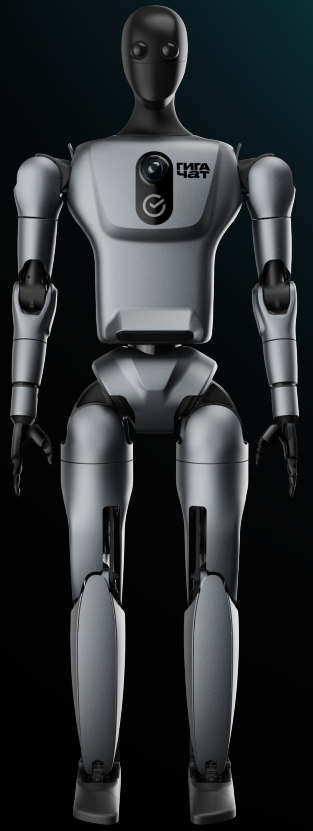
Unitree G1



BD New Atlas



Figure 03



Грин

Что HW сегодня позволяет

✓

Ходить

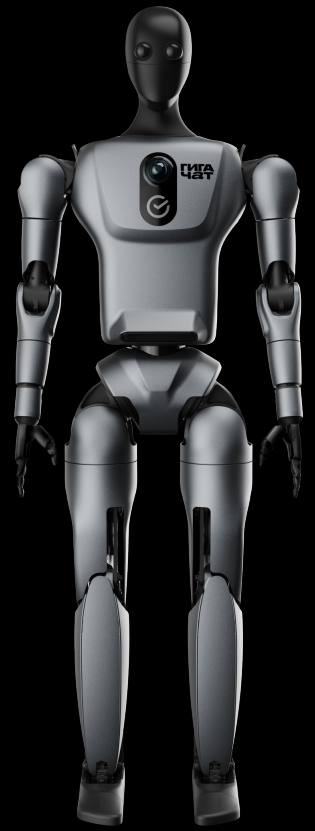
✓

Танцевать/ динамичные движения

✓

Простые операции руками

HW платформы будут разные



Антропоморфная платформа



Колесная платформа с манипуляторами



Колесно-шагающие платформы

????

... и другие формы, оптимальные под условия и сценарии*

*Какой будет оптимум для каких юзкейсов – еще предстоит выяснить

Для массового распространения нужен ~~простой советский~~:

01 **Ресурс приводов**

Сотни → тысячи часов до отказа

---

02 **Тактильное / силомоментное очувствление**

Сейчас только зрение → шнурки вслепую не завязать

---

03 **Коллаборативность by design**

Безопасная работа рядом с человеком из коробки
Сейчас — клетка или kill-switch


Camera 01

Задаем движение: Motion Control

Можем крутить моторами, а хотелось бы:

Стоять
удерживать центр масс

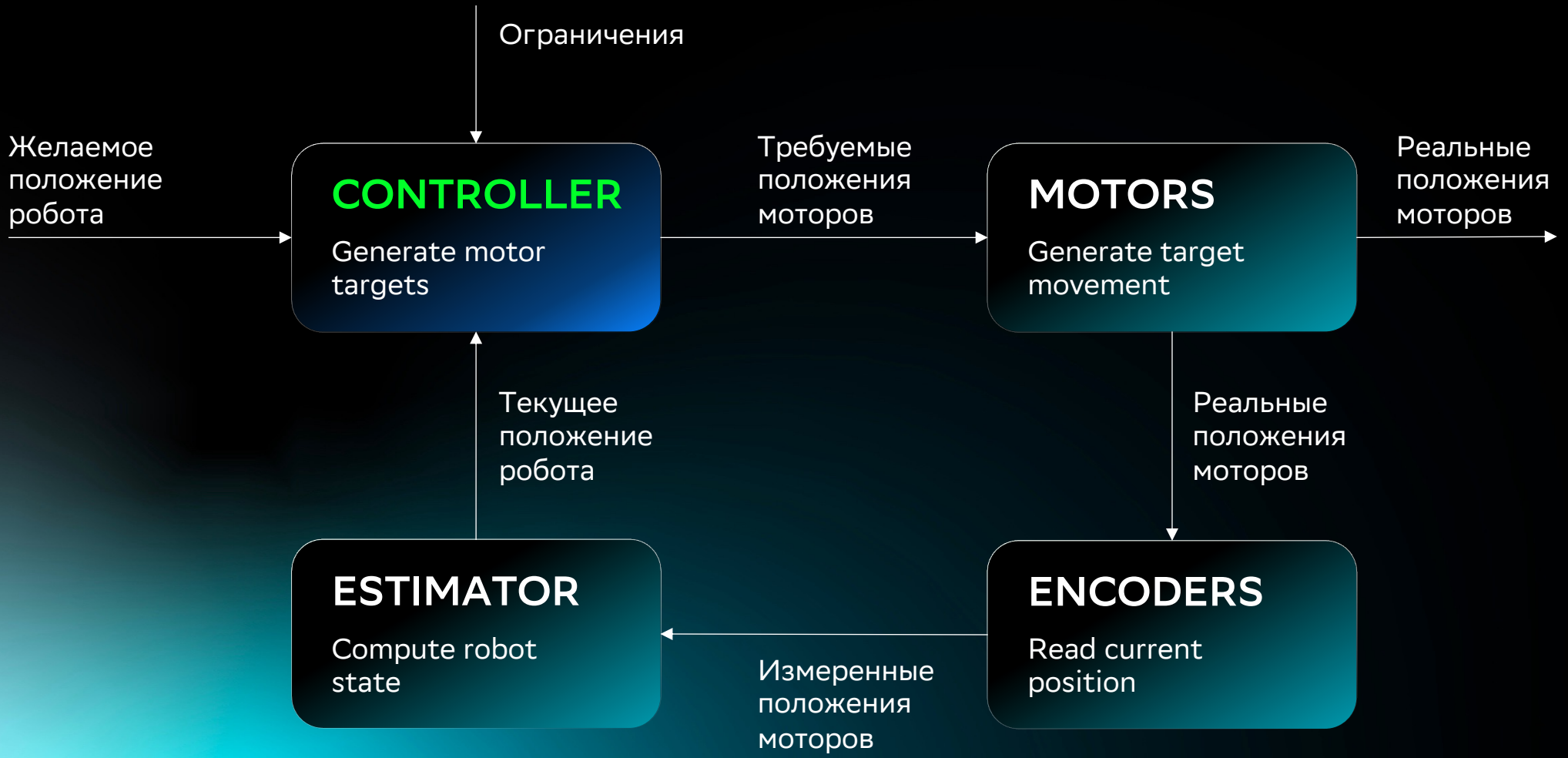
Работать
двигать руками и ногами согласованно

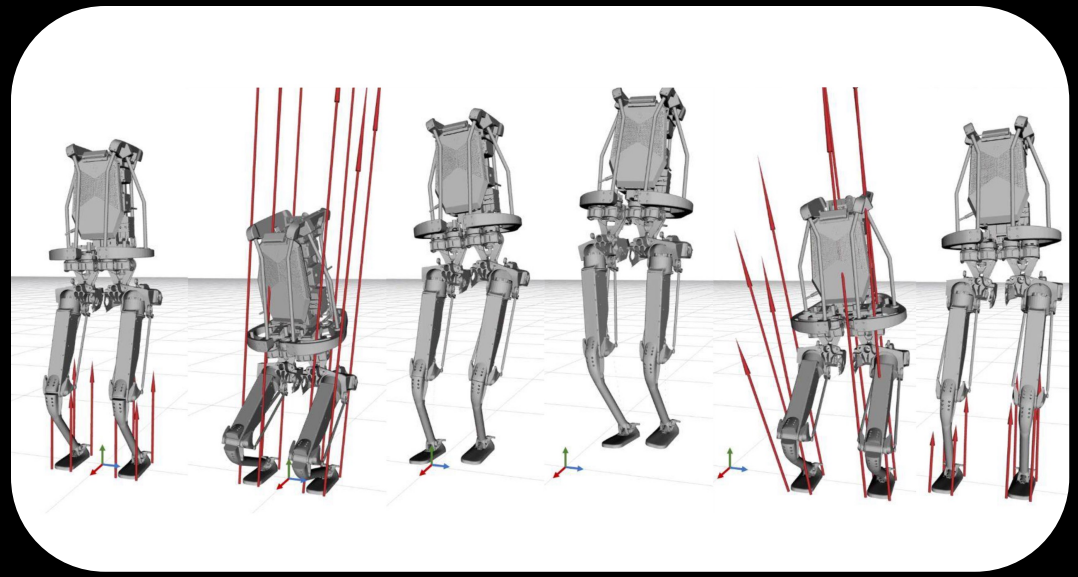
Ходить
следить за переносом ног

Бегать / танцевать
амплитуды + изящество

Для этого придумали Motion Control

Интерфейс наружу: Желаемые положения робота в времени – траектории в joint/cartesian space







MPC

Явная модель динамики
+ оптимизатор в реалтайме

Требует точную модель динамики робота
Более управляем
Надо прогать под каждую задачу отдельно
Плохо параллелится - нужно хорошо оптимизировать под CPU

RL

Нейросеть policy, обучена в симе
на миллиардах эпизодов

Позволяет генерализировать движения
Больше искусство, чем инженерия
Не можем точно управлять деталями движений

В 2024–2025
RL стал primary

И позволяет делать
всякое





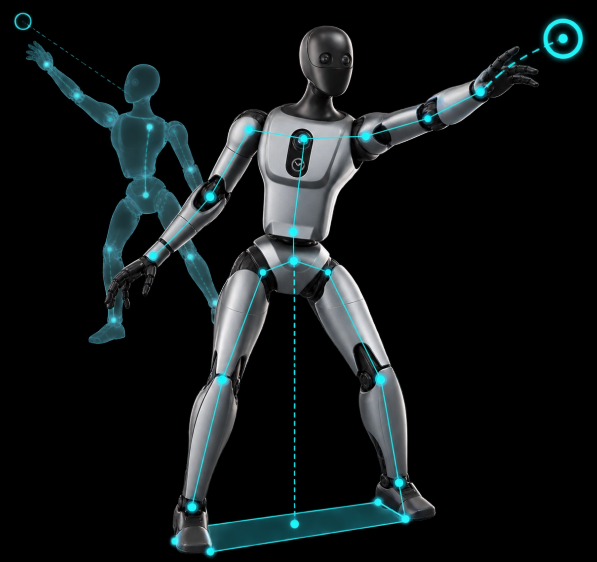


RL политики бывают разные

WBC Generalist

Воспроизводит и балансирует движения всего тела робота

Используется для Teleop и VLA



WBC Specialist

Заточен под определенные движения

Используется для танцев, динамических движений



Walk policies

Делают немного motion planning - планируют двигать ногами и руками для ходьбы



бОльшая часть вирусных видео – это HW + Motion Control

Искусственный «искусственный интеллект»: Оператор с пультом

Команды/вызов записанных движений

LAYER 2 · MOTION CONTROL: RL WBC

Положения моторов

LAYER 1 · HARDWARE · Приводы · сенсоры · батареи · HAL



Учим глаголы: VLA

Можем стоять, ходить, двигать руками, а надо:

Брать/класть
подвести руку, схватить

Упаковывать
поместить предмет внутрь, закрыть упаковку

Переносить
держать рукам и идти

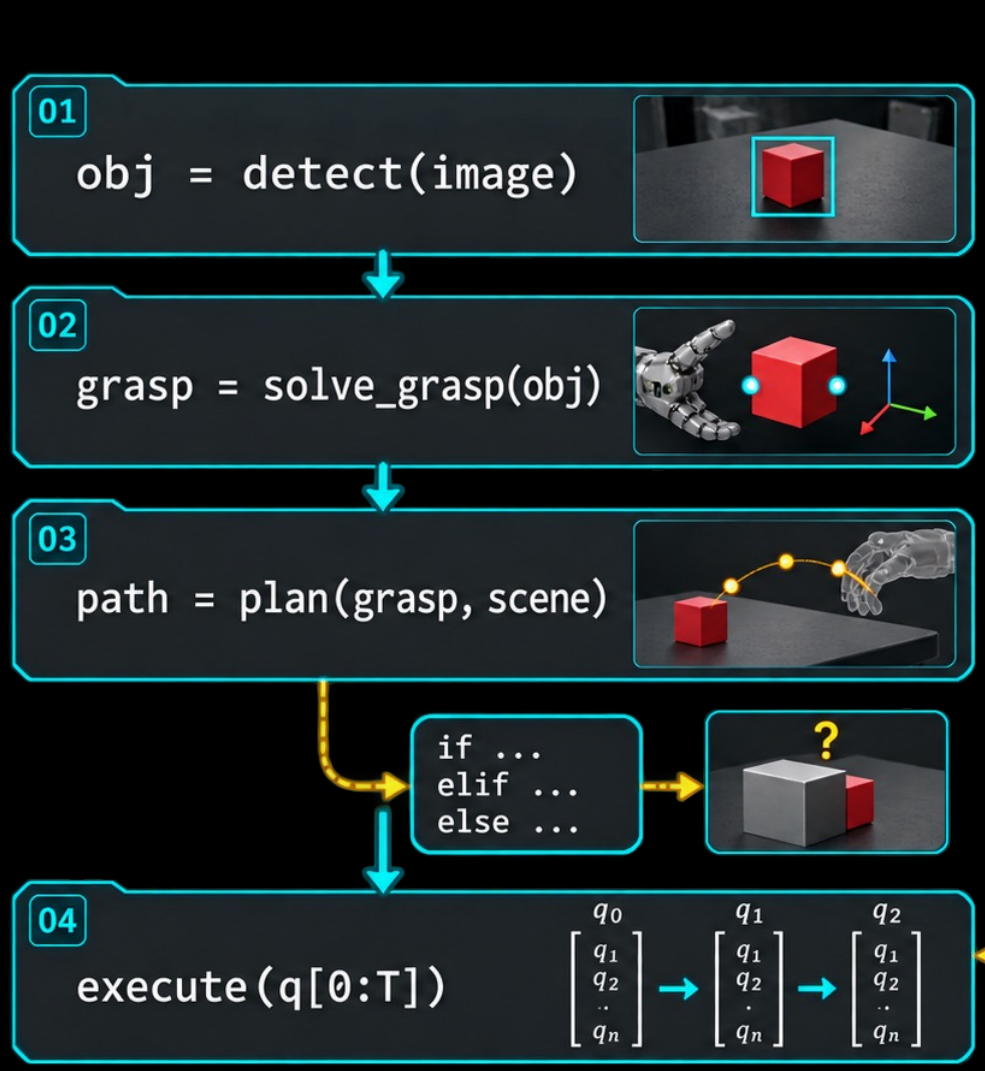
Бегать / танцевать
амплитуды + изящество

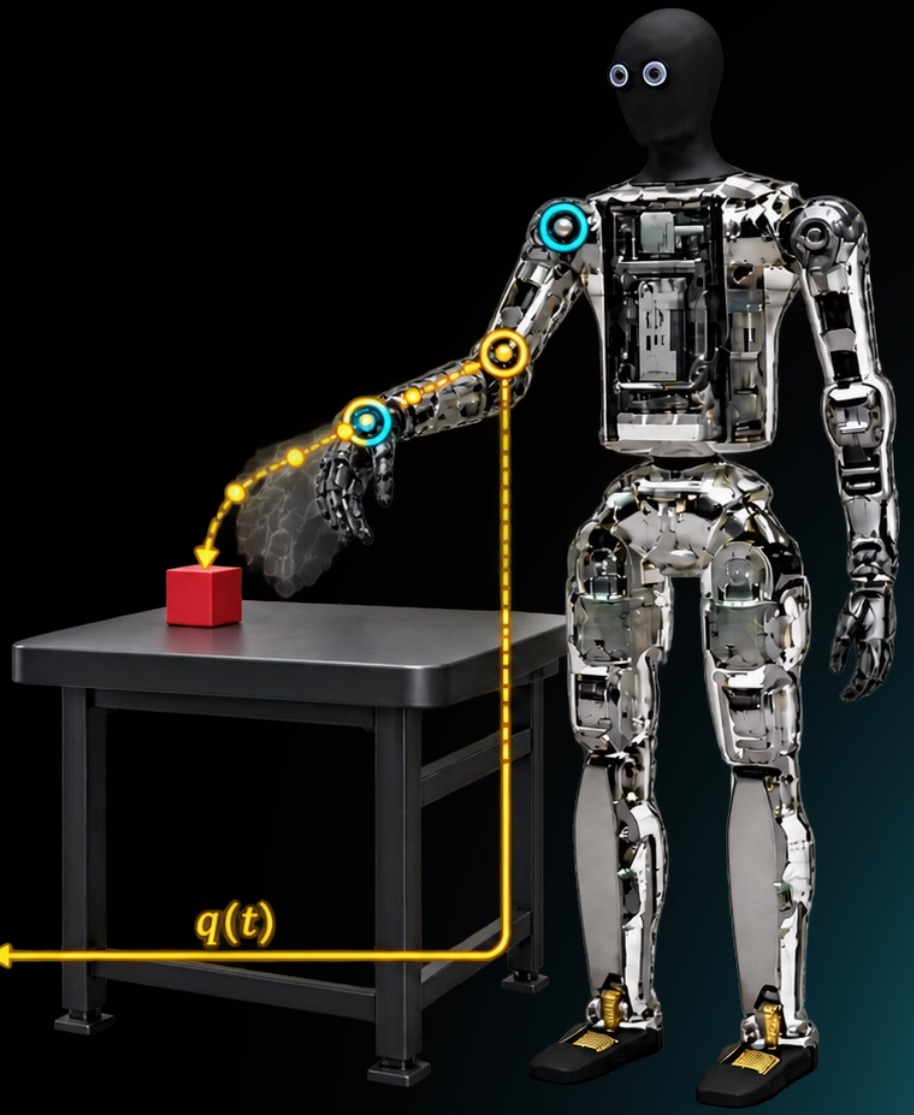
Нужно соединить восприятие, смыслы и движение



Модель ДОЛЖНА сопоставлять **сенсор → смысл → движение**. Потоки должны быть **синхронизированы во времени** — без выравнивания нет связки.

Раньше мы хардкодили алгоритмы действий



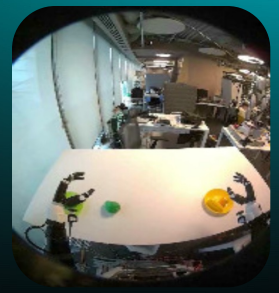


Сегодня для этого есть VLA

”Pick coke can”

”clean up”

”do something”







Joint1 0.75
Joint2 0.21
Joint3 1.17
....
Joint32 0.52

Vision Language Model

Action expert

Joint1 0.75
Joint2 0.21
Joint3 1.17
....
Joint32 0.52

Пять шагов обучения Green-VLA

Метафора: как вырастить повара — от эрудита до шефа

БЕЗ РОБОТА: КАРТИНКИ И ТЕКСТ

С РОБОТОМ: ЗАПИСИ ДВИЖЕНИЙ И СВОИ ПОПЫТКИ

L0

Базовая модель

Учим: видеть картинку и говорить о ней

> Эрудит, который ни разу не держал нож

Готовая модель «картинки + текст»

→

L1

Физический мир

Учим: где что лежит, за что браться, куда тянуться

> Смотрит глазами повара: где ручка у кастрюли

~24 млн примеров из интернета

→

R0

Общая практика

Учим: базовые движения руками на разных роботах

> Стажёр смотрит, как работают другие, и повторяет

3000+ часов записей движений

→

R1

Своё тело

Учим: работать в теле конкретного робота

> Выходит на свою кухню: свои ножи, своя плита

Дообучение на данных этого робота

→

R2

Свой опыт

Учим: доводить дело до конца, исправлять ошибки

> Готовит сам, шеф говорит «вышло» или «нет»

Обучение на собственных попытках

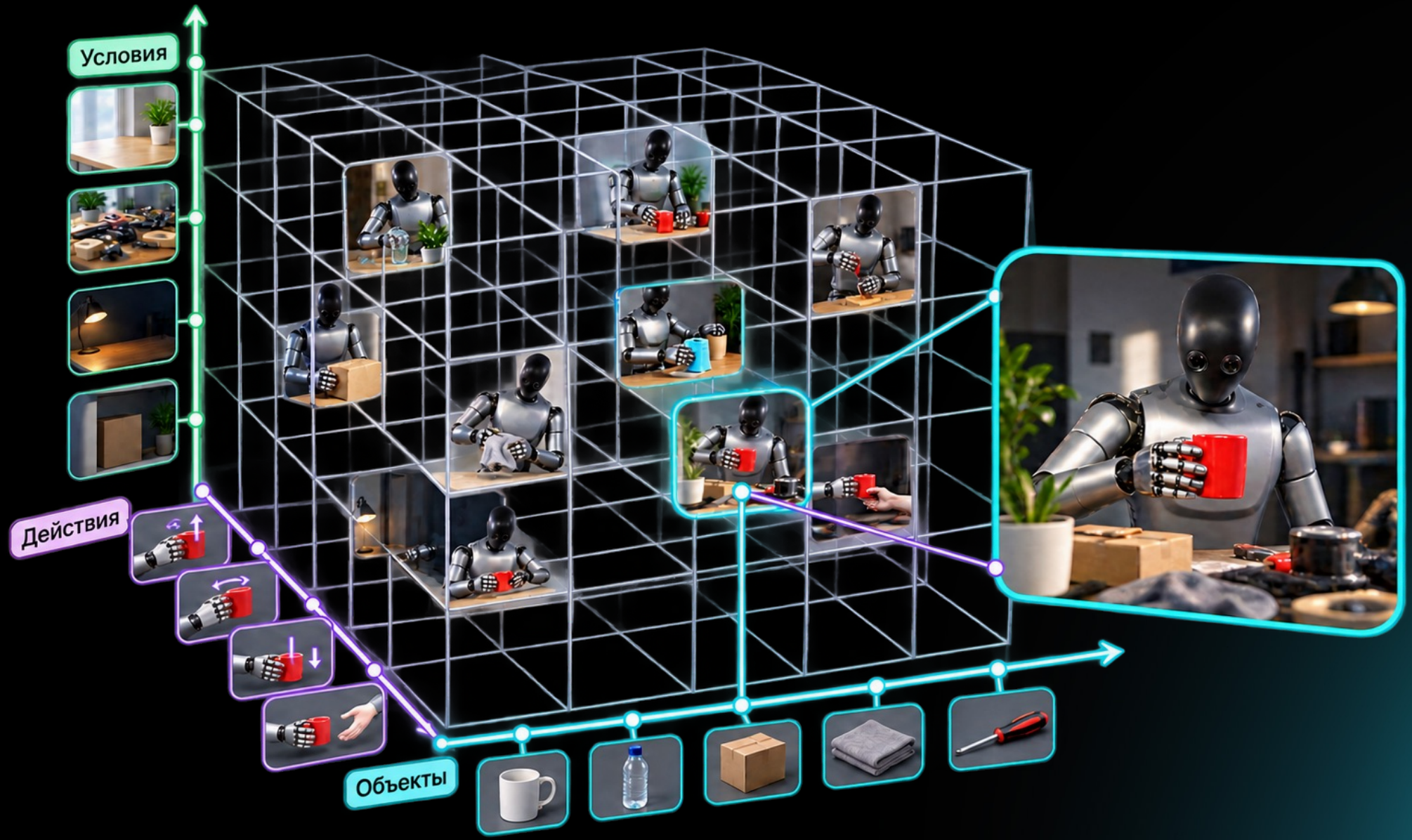
Успех на тест-стенде WidowX (симулятор)	45%	55%	79%

Источник: Green-VLA, arXiv 2602.00919

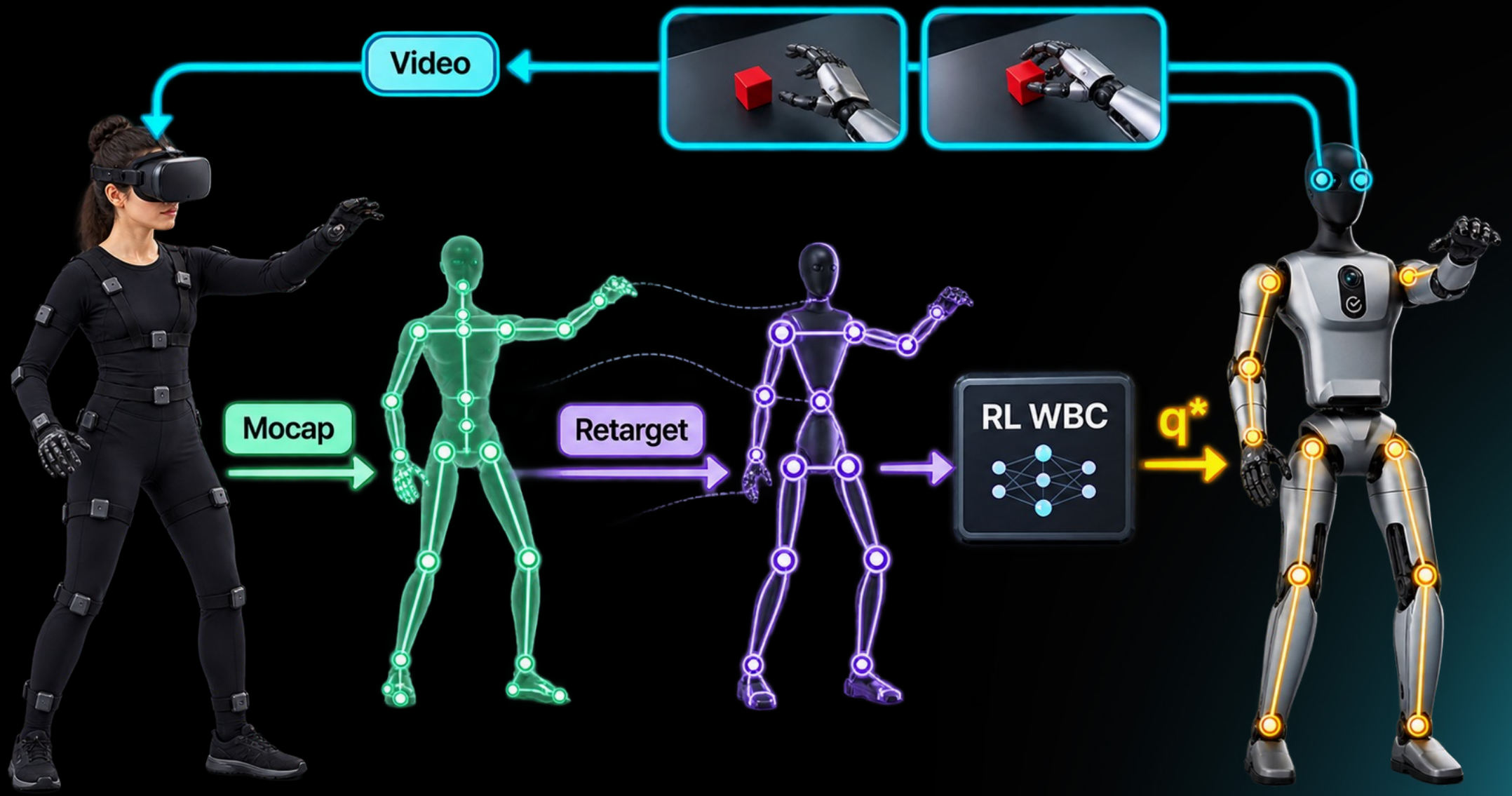

禁止吸烟
No Smoking
AIJ

Данных нужно много и разных

Данных нужно много и разных



Самый очевидный - телеуправление







Можно собирать данные с человеков



Можно собирать данные с человеков

Head ZED



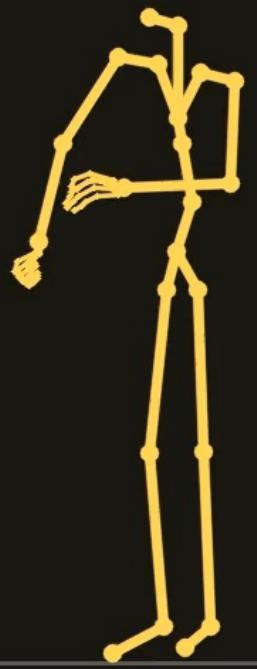
Left wrist ZED



Right wrist ZED



PICO skeleton + hands



Green v3.0



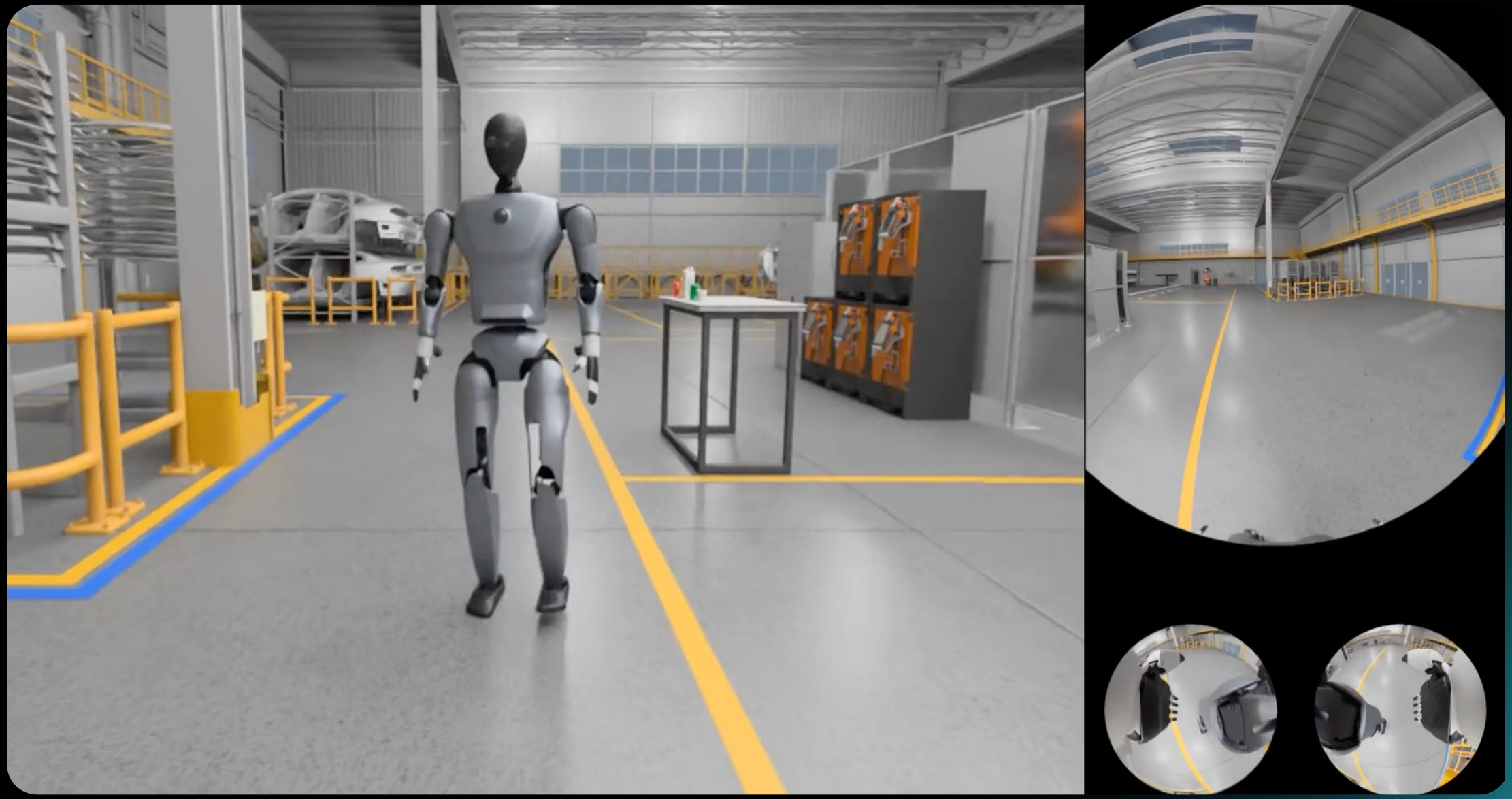
Left hand



Right hand



И из симуляторов



Все* умеем, нужен план

~~(надежный как швейцарские часы)~~

*Ну конечно же далеко не все, глаголов тысячи, объектов миллионы, локаций миллионы, не везде success rate высокий

Учимся планировать: High Level Reasoning

Доминирующая архитектура — иерархия

SYSTEM 2 · SLOW · 1–10 HZ

Большая VLM/LLM

декомпозирует задачу · реагирует на ошибки → переплан

«возьми чашку», «налей воду»

continuous language space, не function calls

SYSTEM 1 · FAST · 10–50 HZ

VLA

исполняет инструкцию end-to-end, генерирует положения робота

Модели из коробки даже в какой-то мере это умеют

ты - cognition робота. Пользователь дал тебе задачу "Наведи порядок на кухне", ты камерами видишь вот такую кухню (картинка).
Составь набор вызовов для VLA, чтобы навести порядок.

Программа вызовов

```
# V = вызов VLA. Инструкция уходит как есть, на английском (язык обучения политики)
# [C] = работа самого cognition: восприятие, решения, проверки
# ✓ = результат, который cognition проверяет по свежему кадру

── 0 · Состояние ──────────────────────────────────────────
[C] осмотреться: мойка, посудомойка, индикаторы плиты (ON / «H» остаточный нагрев)
[C] задать 3 вопроса

── 1 · Пол (заодно безопасность самого робота) ────────────
V01 pick up the pink cloth from the floor                    ✓ на полу нет
V02 put the pink cloth in the laundry basket       # места нет в карте → у двери, в отчёт
[C] рассмотреть тёмный предмет у духовки → мусор или вещь
V03 pick up the dark object from the floor near the oven
V04 put it in the trash bin under the sink | on the counter by the fridge

── 2 · Еда (только при ответе «не готовлю») ───────────────
V05 open the refrigerator door                               ✓ записка на месте
V06 put the plate with food on an empty shelf in the refrigerator
V07 close the refrigerator door                              ✓ закрыта плотно
# сковорода остаётся на плите под крышкой, судьбу еды решает человек

── 3 · Нож отдельно, грязное к мойке ──────────────────────
V08 pick up the knife from the cutting board by the handle
V09 place the knife flat behind the sink, blade facing the wall
V10 put the two mugs from the corner counter next to the sink
V11 put the ladle next to the sink

── 4 · Мытьё ──────────────────────────────────────────────
[C] сушилка занята → нужно место для сушки
V12 spread the gray towel flat on the counter next to the dish rack
```



Что работает

- ✓ Декомпозиция коротких и средних задач
- ✓ Recovery на простых ошибках
- ✓ Команды на естественном языке

Что не работает

- Long-horizon с неопределённостью (>5–10 шагов)
- ~~Common sense про физику~~
- Memory & continual learning
- Многоагентная координация
- Стоимость инференса



Робот = 4 слоя

↓ задача от пользователя

LAYER 4 · HIGH LEVEL REASONING: VLM

✓ Короткий план

! long-horizon, memory

↓ инструкции на естественном языке, 1-10 HZ

LAYER 3 · MOTION PLANNING: VLA

✓ Pick&place · simple manipulation

! generalization gap

↓ желаемые положения робота, ~20–50 Hz

LAYER 2 · MOTION CONTROL: RL WBC

✓ RL неплохо воспроизводит траектории

! сети-специалисты, слепые контроллеры

↓ моменты в моторах, 500Hz

LAYER 1 · HARDWARE · Приводы · сенсоры · батареи · HAL

✓ Ходьба · манипуляции · паркур

! ресурс · тактильность · коллаборативность

Так когда можно будет купить себе?

Кто следил за LLM в 2019?

Помните GPT-2?

В 2019 об LLM говорить было не принято

GPT-2

2019

«смешной автокомплит»

« нет настоящего интеллекта »

« галлюцинирует »

« никакой агентности »

7 ЛЕТ

~~GPT 5.5 / Opus 4.7~~ GPT 6.1 / Opus 5.5

2026

Claude ~~4.7~~ 5.5 пишет код часами автономно

GPT ~~5.5~~ 6.1 решает PhD-задачи через цепочки

Агенты с computer use ходят по браузеру

Триллионы $ AI-капитализации, миллиарды пользователей

МОДЕЛИ БЫСТРО УЛУЧШАЮТСЯ В 3D REASONING

Situational awareness

https://situational-awareness.ai/
Leopold Aschenbrenner, 2024

+10^5

Количество эффективного компьюта на обучение SOTA моделей

По мотивам Situational Awareness · Aschenbrenner

Что разогнало LLM

LLM · 3 МНОЖИТЕЛЯ

Compute
на порядки больше железа в трейне

Algorithmic efficiency
те же возможности дешевле, год за годом

Unhobbling
длинный контекст, тулы, мультимодальность, ризонинг, агенты

Когда множители работают одновременно — прогресс в **OOMs**, не процентах.

по мотивам Situational Awareness · L. Aschenbrenner

Что разогнало LLM — и что разгоняет robotics

LLM · 3 МНОЖИТЕЛЯ

Compute
на порядки больше железа в трейне

Algorithmic efficiency
те же возможности дешевле, год за годом

Unhobbling
длинный контекст, тулы, мультимодальность, ризонинг, агенты

+1

ROBOTICS · 4 МНОЖИТЕЛЯ

Данные ★
телеоп + синтетика + видео + флоты

Компьют
модели и FLOPs обучения

Алгоритмика
архитектуры, training, sample efficiency

Unhobbling
тактильное, long-horizon, continual learning

Когда множители работают одновременно — прогресс в **OOMs**, не процентах.

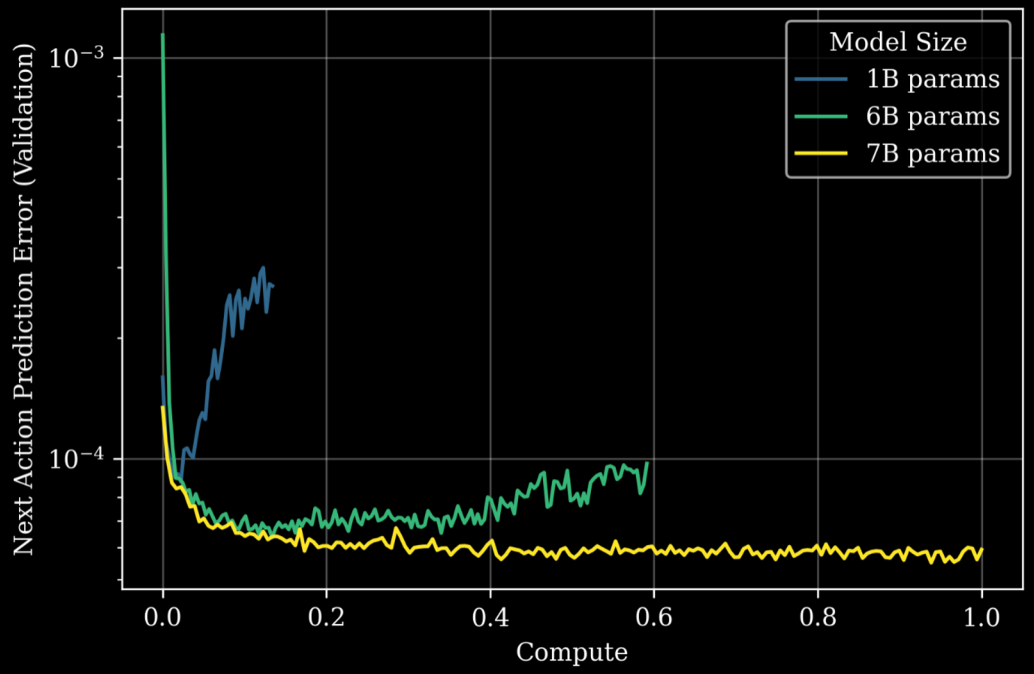
по мотивам Situational Awareness · L. Aschenbrenner

Unhobbling — чем бустим перформанс

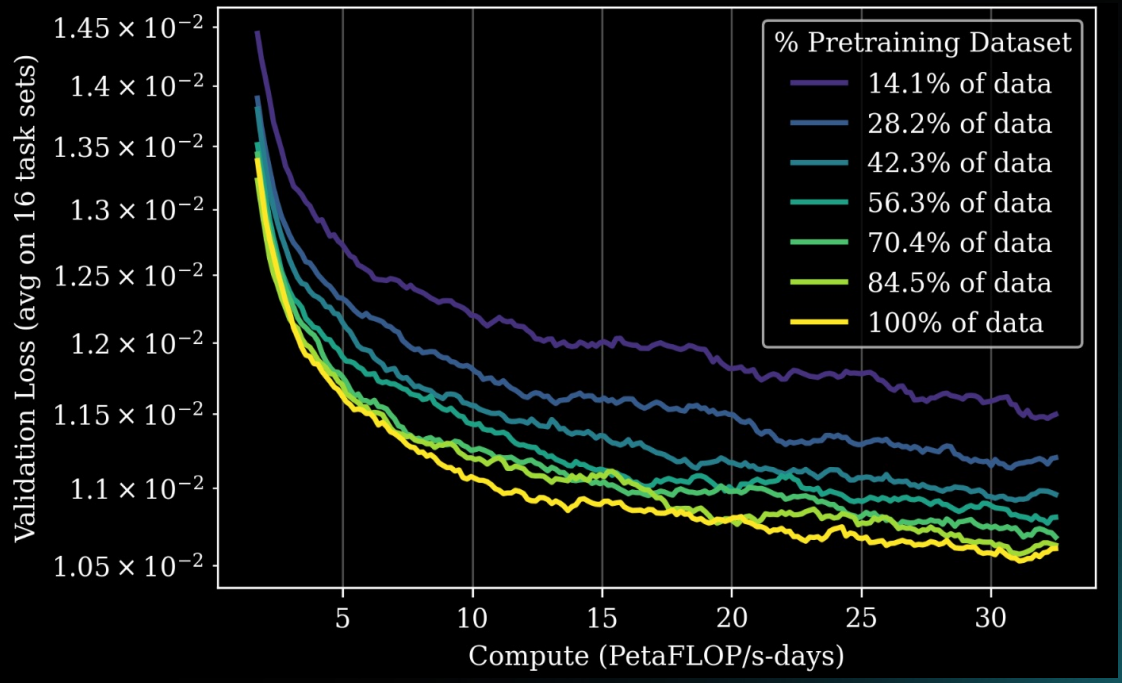
01	**Continual learning из деплоя**	Self supervising RL Finetune
02	**World Models**	Предсказание состояния мира и генерация движения
03	**Long-horizon memory**	EchoVLA · RMBench
04	**Тактильное / силомоментное**	ForceVLA · Tactile-VLA · HapticVLA · 2025–2026
05	**Большая reasoning-модель сверху**	Растёт сама — бесплатный апгрейд

Scaling Laws подтверждаются

Prediction Error от размера модели



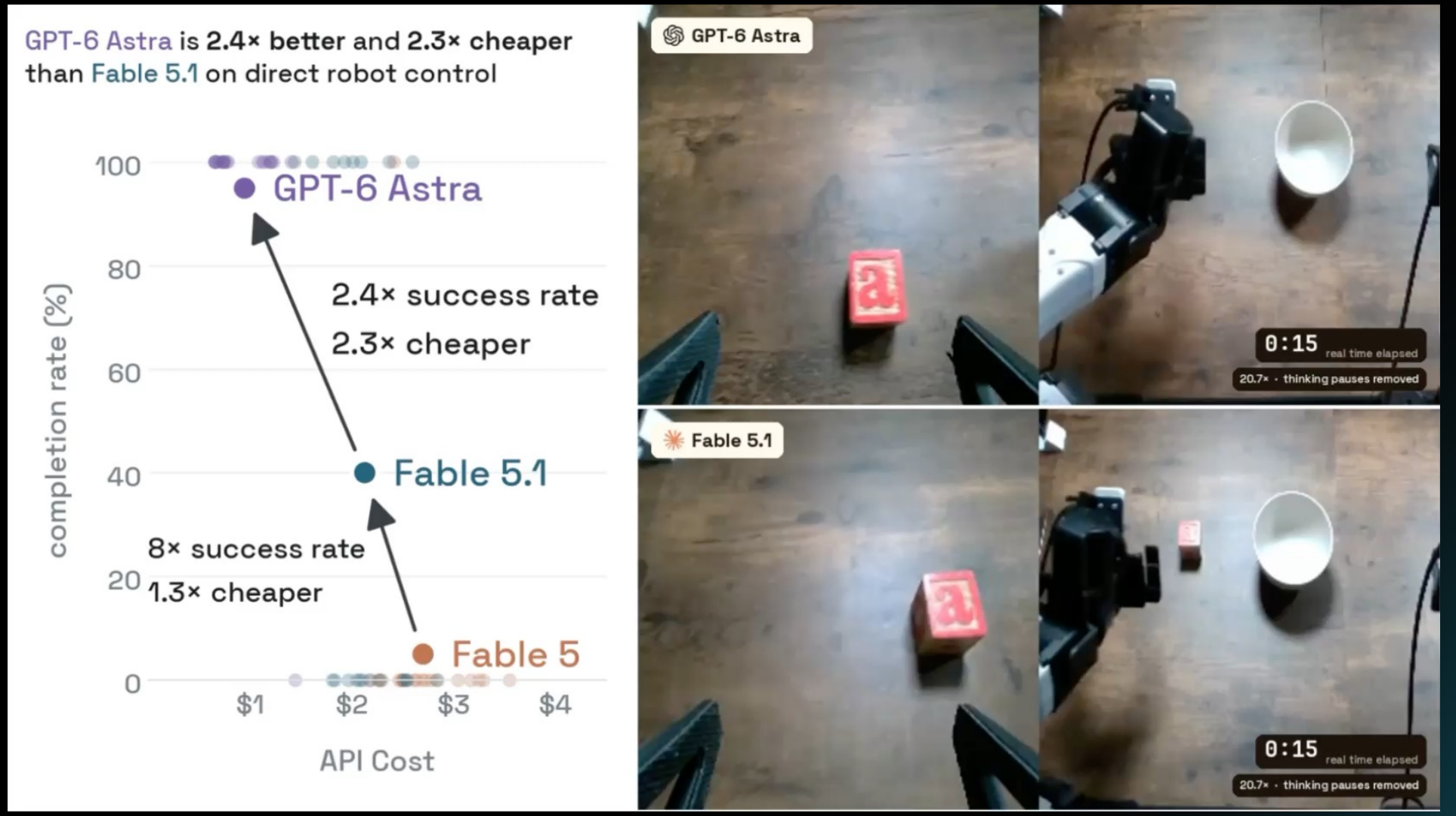
Val Loss от размера pretrain датасета



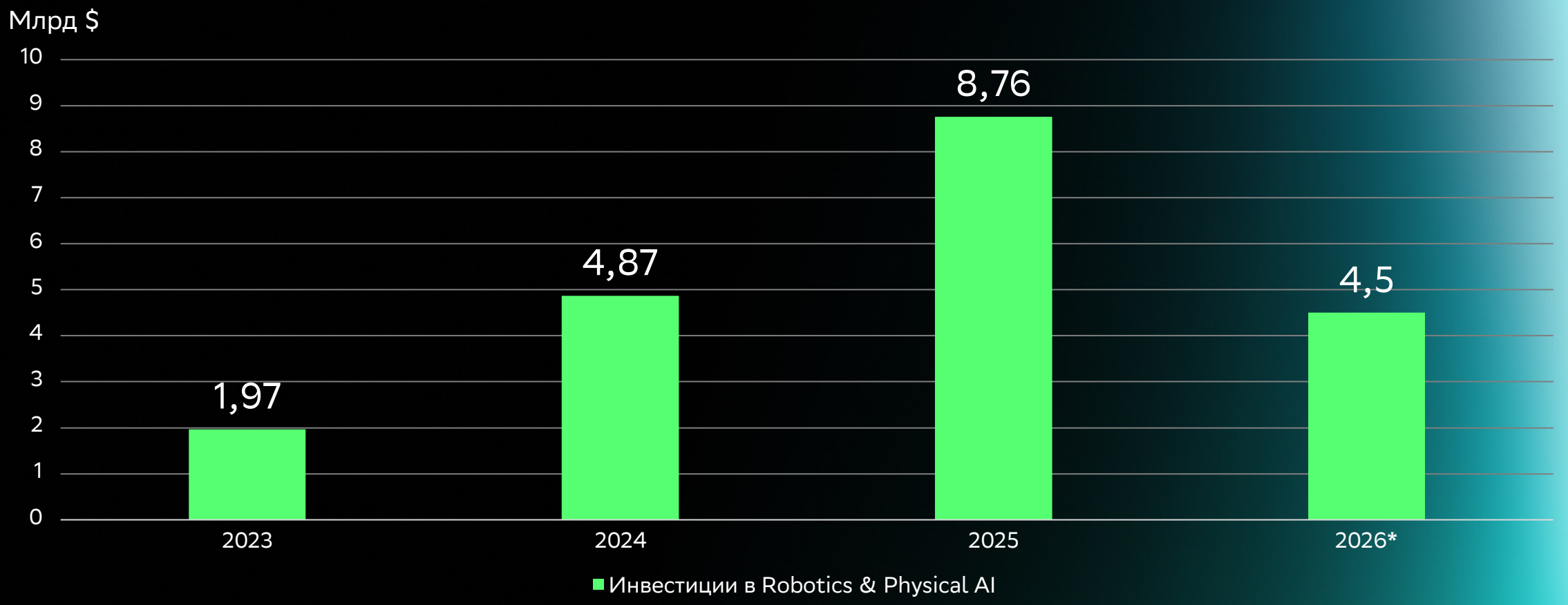
Kaplan et al. опубликовали scaling laws для LLM в 2020 — через 2 года после GPT-2. Мы сейчас в этой точке для роботов.

https://generalistai.com/blog/nov-04-2025-GEN-0

Фронтирные модели из коробки начинают понимать, как управлять роботами



Ну и как прокси метрика – инвестиции разгоняются



Мы приближаемся к GPT-3-моменту



2024–2025

GPT-2 момент

базовая генерализация, демо

2026–2028

GPT-3 момент

scale-up данных и моделей · первые широкомасштабные деплои

2028–2030

ChatGPT / RLHF

массовый B2B-продукт

2030–2032

GPT-4 момент

foundation models с broad capability · домашние пилоты

2032+

Claude 4 / GPT-5.5

mature reasoning-агенты с computer use в физическом мире

Могу ошибиться на ±2 года ☺

2026 для роботов

= ~~2019~~ 2021 для LLM

Лет через семь проверим


НАС ВСЕХ ЗАМЕНИТ AI
ТОГДА МЫ ПОЙДЕМ РАБОТАТЬ СВАРЩИКАМИ, ДА?
ДА?
imgflip.com

Центр Робототехники

Профессионалы всех направлений PhysicalAI



Портал о нас + отклики

2017	2019	2023	2025	Сейчас
Создание Лаборатории робототехники Сбера	Лаборатория → Центр	Решение о разработке антропоморфных роботов	Презентация Грина на AIJ'25	Развитие PhysicalAI

Присоединяйтесь к команде!



TG: svyatoslavdm



Портал о нас + отклики