

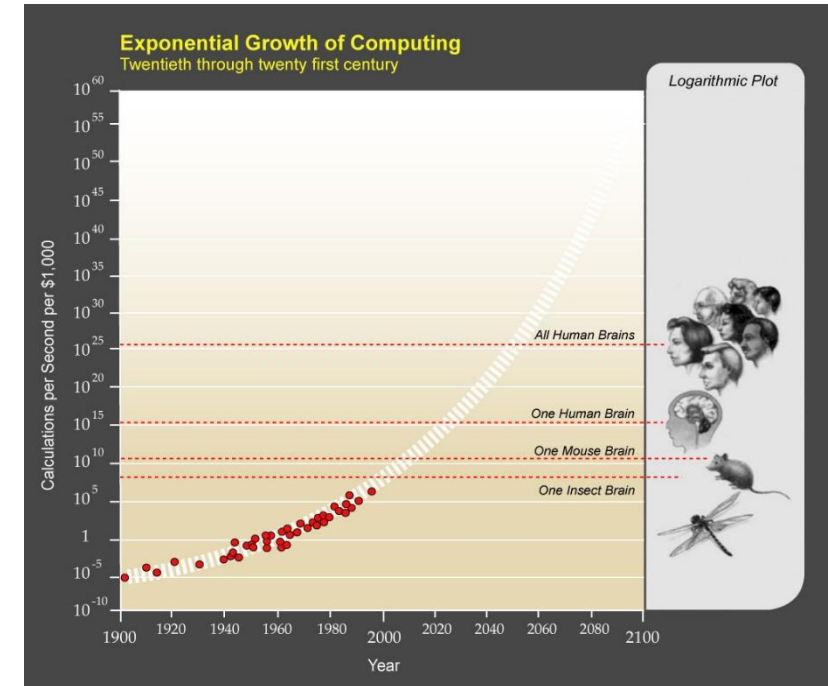
- 
- Как технологии ИИ меняют мир?
 - А как бы нам хотелось?

К.В.Воронцов, завлаб МОСА

От закона Мура к предсказанию сингулярности

«В будущем ИИ...

- ... станет непонятен для людей»
- ... заменит людей, лишит нас работы»
- ... приведёт людей к праздности и деградации»
- ... послужит для узурпации власти над миром»
- ... станет автономным и выйдет из-под контроля»
- ... уничтожит человеческую цивилизацию»
- ... и всю биологическую жизнь на Земле»
- ... продолжит вместо нас эволюцию разума»
- ... захватит Землю и устремится в космос»
- ...



**Нам повезло заниматься ИИ в тот момент, когда
эта технология имеет наглость менять мир**

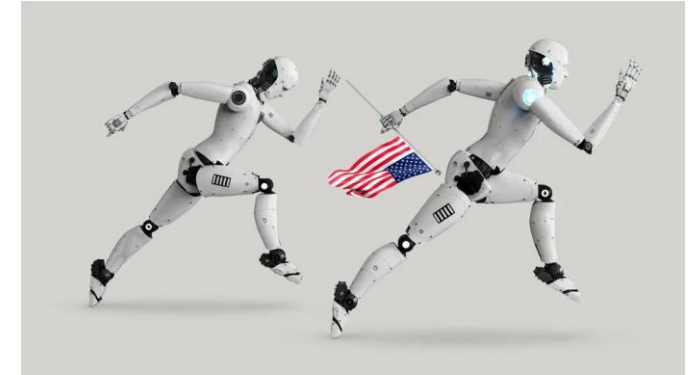
Этот факт начинает менять нас самих, потому что
мы уже не можем отмахнуться от экзистенциальных вопросов,
каждый из которых начинается с «*Чёрт возьми*»...

Чёрт возьми,

- мы прогнозируем будущее или конструируем его?
- мы меняем мир или просто пишем очередную статью?
- разве можно стать лидерами в ИИ, не ставя перед собой задачу изменить мир?
- что мы хотим изменить в мире с помощью ИИ?
- у нас есть внятные образы будущего?
- какие постановки задач в ИИ способны менять мир?
- как согласовать краткосрочные задачи с долгосрочными стратегиями?
- могут ли менять мир те, у кого нет 1B на обучение больших моделей?
- чем заниматься в ИИ — создавать возможности или устранять угрозы?
- разве есть реальные угрозы, а не пустые пугалки для обывателей?

Как перестать догонять и пойти на опережение

Только читать статьи и делать улучшения — означает всегда плестись в хвосте, отставая в исследованиях на 1-2 года.



- Чтобы не отставать, достаточно делать аналогии.
- Чтобы обгонять, нужно задавать новые **свои** цели, смыслы, концепции, ставить и решать **свои** задачи, идущие от собственных потребностей

Что для этого нужно:

системный подход, целеполагание, идеология, фантазия, образы будущего

- **Жизнь человека — это регистрируемый текстовый поток:**

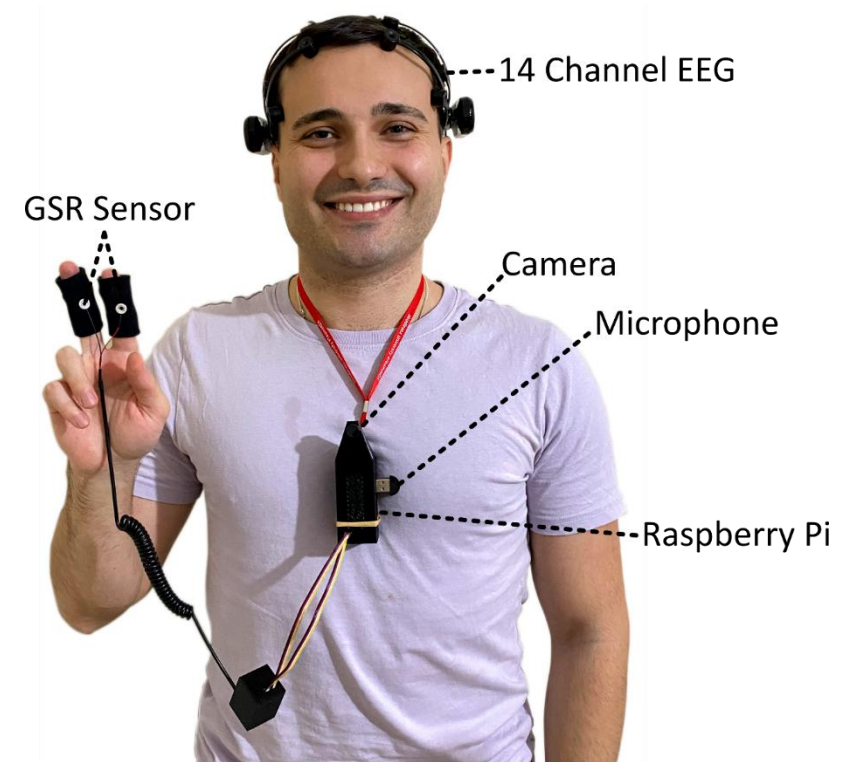
браузеры, почта, соцсети, мессенджеры, пользовательские документы, файлы, системы учёта времени и ведения проектов, ВКС, видео, аудио, голос, VR/AR, ...

- **FPFM (First-person Foundation Model):**

каждый человек — генератор уникального потока семантических векторов, по объёму сопоставимого с Интернетом

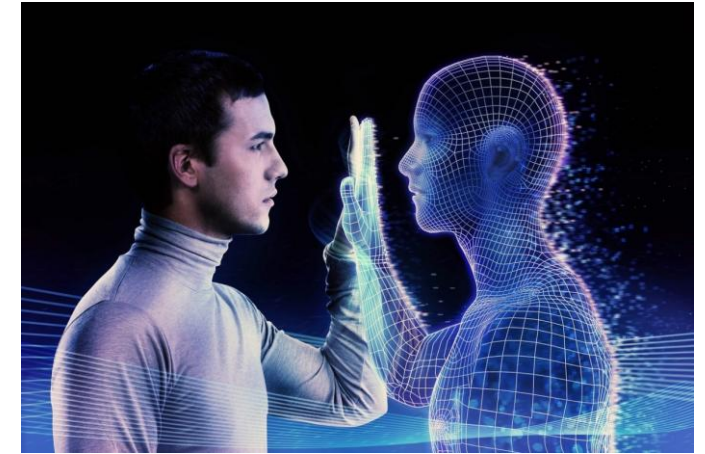
Они жалуются, что у них закончился Интернет?

Но у них остались мы — 8 миллиардов людей



От персонального помощника к цифровом посмертию

- **Пом** становится универсальной записной книжкой и дневником
- **Чел** доверяет **Пому** всё больше своих информационных функций (поиск, обучение, коммуникация, выработка и принятие решений)
- **Пом** перенимает черты личности человека, его *личностный код*
- **Чел** умирает.
Пом становится **Ава**таром, это ценный информационный ресурс
- **Ава** продолжает приносить пользу обществу, выполняя профессиональные и коммуникативные функции **Чела**
- **Ава** отличается от «фабричного ИИ» тем, что обучен на жизни человека, лучше понимает людей, их ценности, цели, чувства, взаимоотношения
- **Ава** продолжает развиваться, накапливая знания и мудрость



Идея: открытая литературная вселенная фантастики ближнего прицела

- **фантастическая гипотеза** — научная, реалистичная, достижимая
- **цель** — обнаружить угрозы и уязвимости развития ИИ на пути к человеко-машинной цивилизации, найти способы их обхода
- **задача** — проверка гипотез о возможности или невозможности, убедительности или неубедительности различных путей развития
- **задача** — попытка «взлома» позитивных сценариев развития, вывода апокалиптического сценария из исходных предположений

Первый литературный конкурс состоялся
в апреле 2025 рамках конференции DataFusion



«Отличительная черта хорошего ученого — умение задавать вопросы, которые до него не задавал никто»



Карл Фристон

«Каждому может повезти,
если при выборе темы
он найдет зарождающееся
направление исследований»



Карл Фристон



Спасибо за внимание!

Воронцов Константин Вячеславович
зав. лаб. МОСА

k.vorontsov@iai.msu.ru

 **Институт
искусственного
интеллекта**
мгу

info@iai.msu.ru

iai.msu.ru

Ломоносовский пр-кт, 27,
корп.1, пом. Г-602
Россия, Москва, 119192