Ежегодная XV международная конференция по научно-технической политике и форсайту, ВШЭ, 14 ноября 2025

Искусственный интеллект: от персональных помощников к цифровому послесмертию

Воронцов Константин Вячеславович

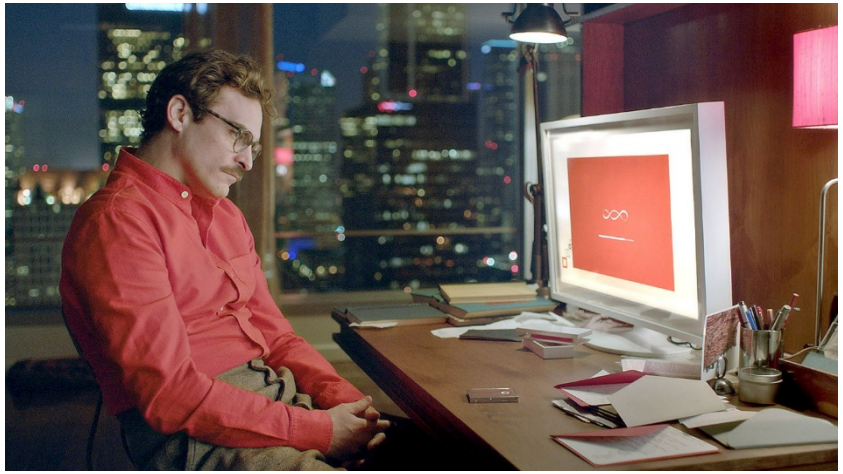
д.ф.-м.н., профессор РАН,
зав. кафедрой математических методов прогнозирования МГУ,
рук. лаб. машинного обучения и семантического анализа Института ИИ МГУ,
зав. кафедрой машинного обучения и цифровой гуманитаристики МФТИ,
зав. кафедрой интеллектуальных систем МФТИ,
г.н.с. ФИЦ «Информатика и управление» РАН

k.vorontsov@iai.msu.ru

Перспектива AI: персональные помощники

Фильм «Она» (Her, 2013)

режиссёр Спайк Джонз, «Оскар» за сценарий



Тот самый парень, который влюбился в операционку
(жанр фильма — фантастическая мелодрама)

- голосовая операционная система
- анализирует всю деловую коммуникацию
- ведёт дела и переписку, генерирует идеи,
- до которых человек не додумался сам
- по контексту ищет информацию в сети
- формирует модель личности пользователя — интересов, привычек,потребностей, деловых и профессиональных компетенций
- понимает эмоции человека, способна манипулировать человеком

Это будущее уже наступило

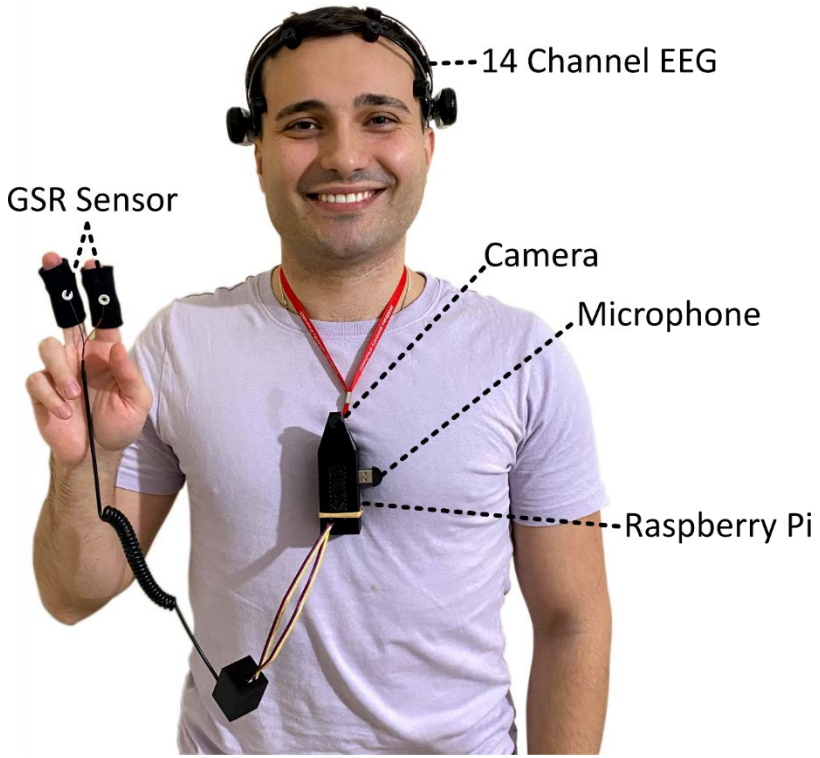
Перспектива AI: следующий шаг

- **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?

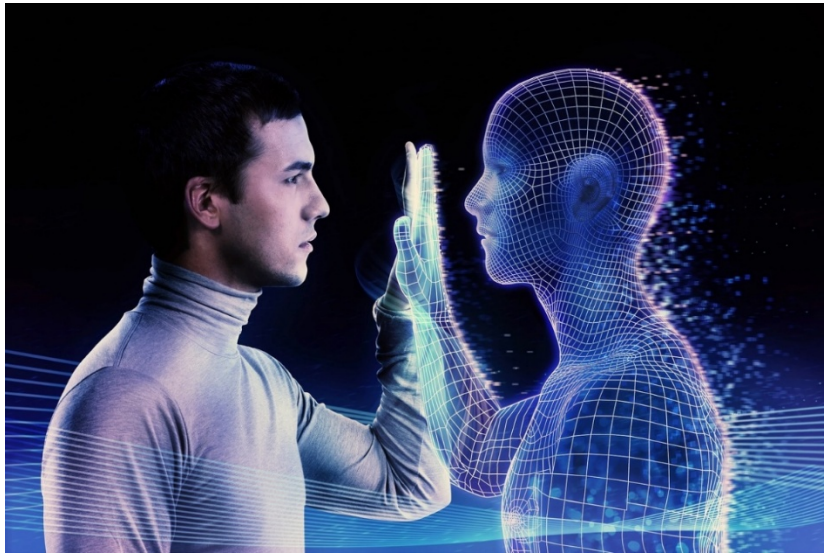
Но у них остались мы — 8 миллиардов людей

D.Barcari, D.Gamez, A.Grig. Recording first-person experiences to build a new type of foundation model. 2024-07-31

Пройдут годы…

- **Чел**овек воспитывает и обучает своего персонального **Пом**ощника всю жизнь, в процессе всей своей разнообразной деятельности
- **Чел** делегирует **Пом**у всё больше своих информационных функций (поиск, обучение, коммуникация, анализ, принятие решений)
- **Чел**+**Пом** в связке накапливает репутацию и социальный капитал
- **Пом** всё лучше замещает **Чел**а, но под его надзором и контролем
- **Пом** становится для **Чел**а записной книжкой и поминутным дневником
- **Пом** перенимает черты личности **Чел**а, его *личностный код*
- **Пом** обладает сверх**чел**овеческими темпом и возможностями развития: вычислительными, поисковыми, коммуникативными, генеративными

Все люди смертны



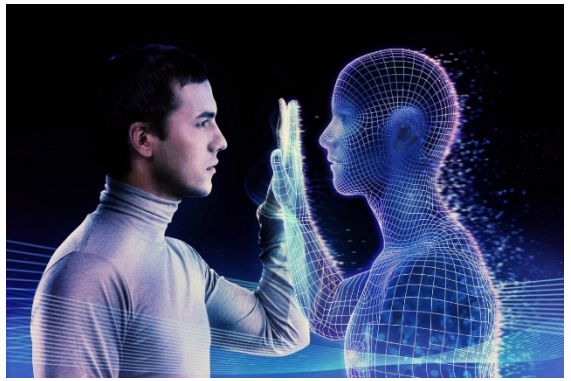
- **Чел** умирает. **Пом** становится **Ава**таром, это ценный информационный ресурс!
- **Ава** продолжает приносить пользу обществу, выполняя профессиональные и коммуникативные функции **ЧелПом**а
- **Ава** переходит в общественное достояние, но не становится полностью автономным, меняются регламенты его эксплуатации
- **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
- **Ава** остаётся доступен для семьи в роли наставника, «хранителя рода»

Послесмертие — не бессмертие, а наследование

- Восприятие и сознание **Чел**овека умирает необратимо со смертью головного мозга
- *Личностный код* передаётся:
  Чел → **Пом** → **Ава**

- **Ава** — это бывший **Чел**, что намного больше «фабрично-безличного ИИ», встраиваемого в машины. Он личность, он мудрее и опытнее людей. Неутомим, неуязвим, наделён естественным аскетизмом.
- **Чел** — это будущий **Ава**, ответственный за качество передаваемого личностного кода, включая репутацию, социальный капитал, знания о человеческой природе, ценностях, приоритетах, целях, задачах.

Аватар, в отличие от Человека



Лишён

- тела, а значит — усталости, боли, лени, тревоги
- потребностей в еде, сексе, отдыхе, гедонизме, лечении, сочувствии, защите от стресса, самовыражении, демонстративности
- стремлений к самосохранению, доминированию, власти
- «внутреннего зверя» — семи смертных грехов (гнева, гордыни, жадности, зависти, уныния, похоти, чревоугодия)

Способен

- управлять роботами, производствами, отраслями, …
- кооперироваться (мгновенно) с другими аватарами для решения трудных для людей задач (в космосе, в океане, под землёй)

Как решать морально-этические проблемы?

- **Чел** может иметь секреты? (*да*)
- **Чел** может прерывать запись потока данных для **Пом** и **Ава**? (*да*)
- **Чел** может устанавливать права доступа к своим данным? (*да*)
- **Чел** должен быть информирован, что общается с **Пом** или **Ава**? (*да*)
- нужно ли «чистилище» при переходе **Пом** → **Ава**? (*да*)
- можно ли клонировать **Ава**? (*нет*)
- всем ли будет доступен **Ава** в режиме чат-бота? на каких условиях?
- кто нанимает **Ава** на работу и как оплачивается его работа?
- с кем и какой секретной информацией может делиться **Ава**?
- как выделяются энергетические и вычислительные ресурсы для **Ава**?
- …

Очеловеченный ИИ: путь к снижению рисков ИИ

О каких рисках развития ИИ мы говорим уже сегодня:

1. утрата людьми контроля над процессами, потеря управляемости систем
2. сверхчеловеческий рост объёмов данных, информации и знаний
3. деградация интеллектуальных способностей людей

Направления R&D, нацеленные на снижение этих рисков:

1. создание ***этики*** и ***идеологии*** человеко-машинной цивилизации
2. создание единого структурированного ***пространства знаний*** (**Пом** общается с **Чел**ом посредством визуализаций, интеллект-карт)
3. создание протоколов ***доверенной*** человеко-машинной коммуникации (**Пом** оставляет **Чел**у целеполагание, ответственность, интерес, развитие)

Какие технологии развивать? (например, AGI)

Идём от целей и задач к технологиям, но не наоборот

1) каковы цели и задачи развития компании, отрасли, цивилизации?
2) какие технологии необходимы и минимально достаточны?
3) для решения каких задач мы собираемся использовать AGI?
4) генераторы текстов и картинок насколько важны для развития?
5) ожидаемые эффекты насколько перевешивают затраты и риски?

ИИ, цифровизация, как и любая технология — это не цель, а средство

Антропоцентричное определение ИИ

Искусственный интеллект —

вычислительные технологии, создаваемые для повышения производительности созидательного интеллектуального труда людей

не замена человека

не «загадочный новый тип разума»

не повод уподобиться Богу и творить «по образу и подобию Своему»

Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,

зав. лабораторией
машинного обучения
и семантического анализа
Института ИИ МГУ,

зав. кафедрой ММП ВМК МГУ,

зав. кафедрой МОЦГ МФТИ,

г.н.с. ФИЦ ИУ РАН

Цивилизационная идеология

ДЗЕН-канал

https://dzen.ru/civideology

k.vorontsov@iai.msu.ru

http://www.MachineLearning.ru/wiki?title=User:Vokov