



Экономический факультет

МГУ имени М.В. Ломоносова



Футурология ИИ: как персональные помощники будут влиять на социально-экономическое развитие

ИИ

МГУ

Воронцов Константин Вячеславович

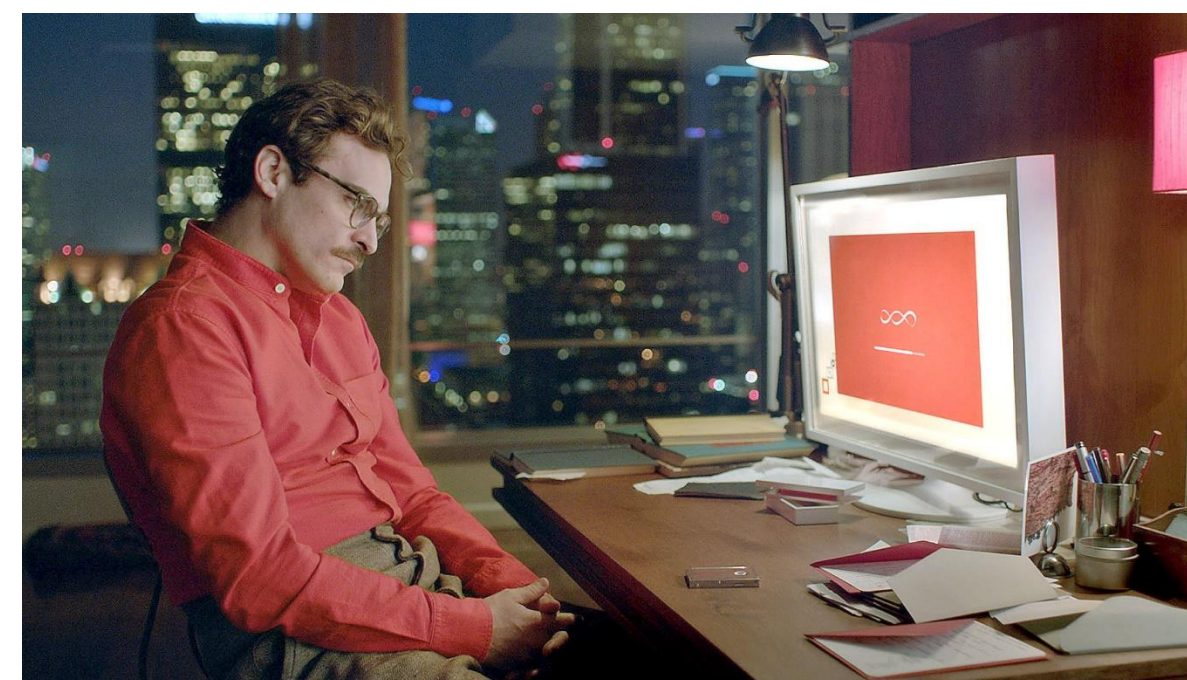
д.ф.-м.н., профессор РАН, руководитель лаборатории
машинного обучения и семантического анализа

Институт искусственного интеллекта МГУ им. М.В. Ломоносова

Возможно ли предсказывать развитие технологий?

Фильм «Она» (Her, 2013),
режиссёр Спайк Джонз, «Оскар» за сценарий

- 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 — его интересов, привычек, потребностей, деловых и профессиональных компетенций
- понимает эмоции человека, способна манипулировать человеком



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

Это будущее уже наступило

Следующий шаг

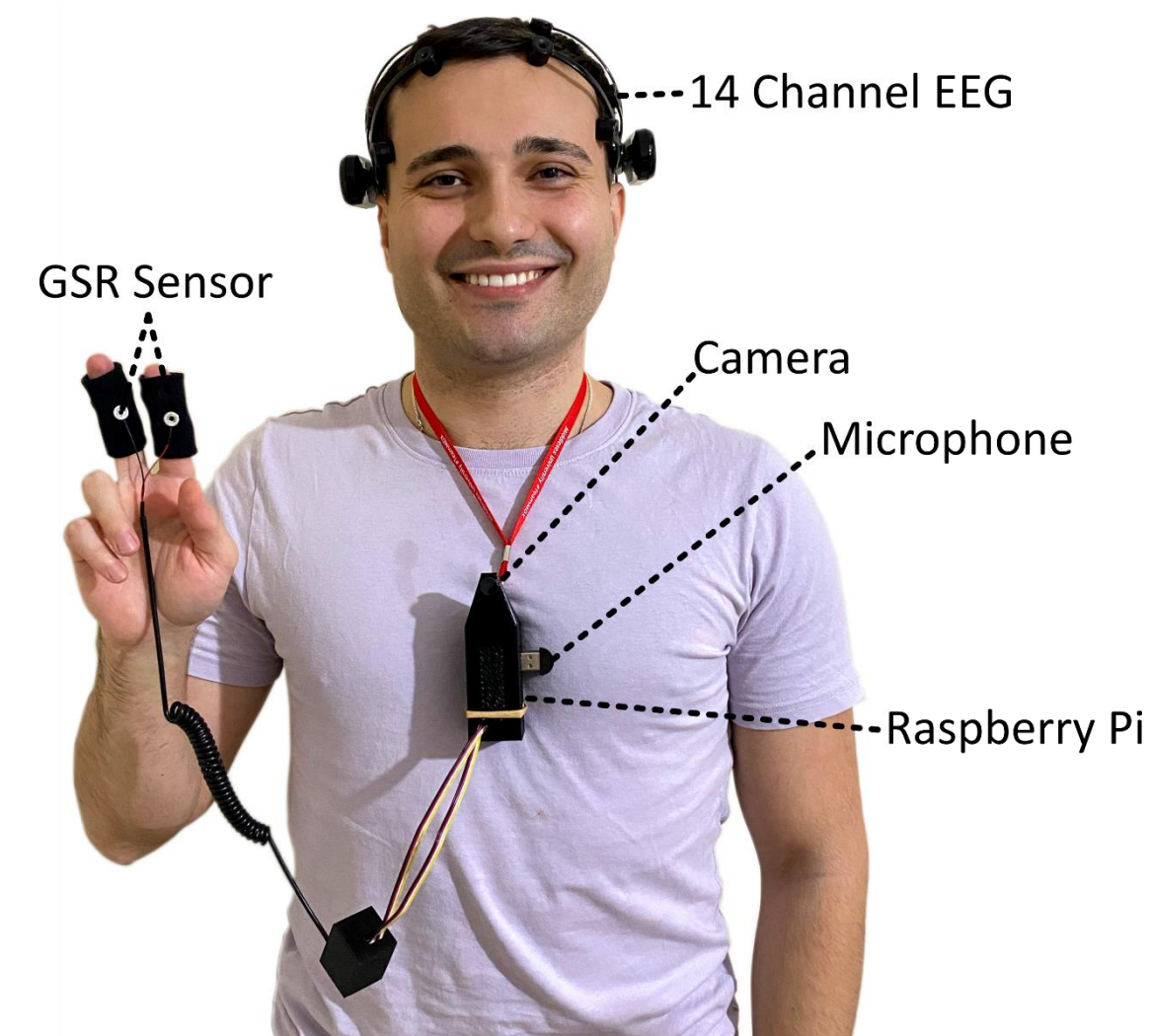
- **Жизнь человека — это регистрируемый текстовый поток:**

браузеры, почта, соцсети, мессенджеры, файлы пользователя, 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



Пройдут годы...

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

ЧелПом: социально-экономические изменения

- интенсификация доступа к знаниям, деловых коммуникаций,
- бизнес-процессов, темпов развития, научно-технического прогресса

Риски

- сверхчеловеческий рост объёмов данных, информации и знаний
- автономизация процессов управления, латентная утрата управления
- изменение рынка труда, переформатирование профессий

Что делать (сейчас, из нетривиального)

- создавать единое структурированное **пространство знаний**,
- протоколы **доверенной** человеко-машинной коммуникации,
- закладывать **идеологию** человеко-машинной цивилизации
- ...

Что такое «знания» и «пространство знаний»



мудрость
(wisdom)

самое главное:
смыслы, ценности, цели, задачи



знания
(knowledge)

информация, структурированная
для удобства понимания и
практического использования



информация
(information)

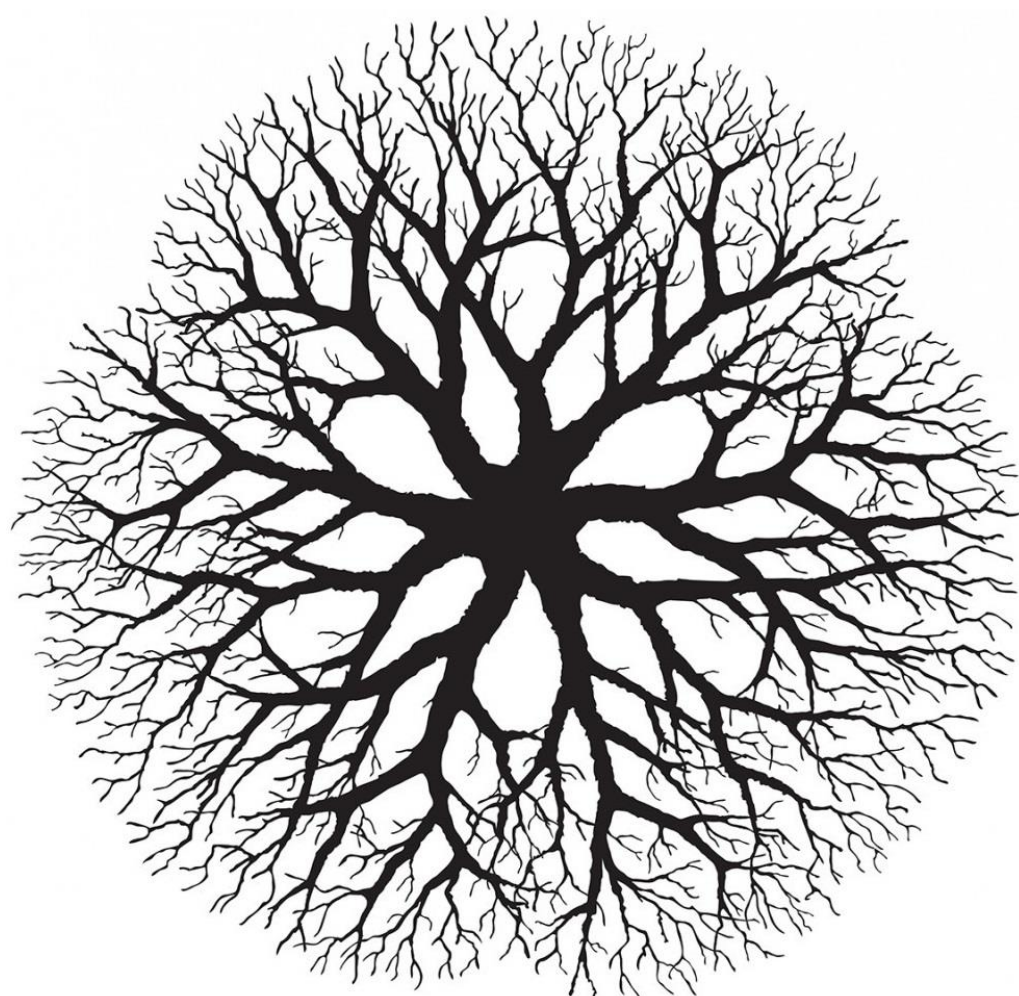
результат обработки и
анализа данных



данные
(data)

зарегистрированные факты
окружающей реальности

Смысловое ядро пространства знаний



естественно-научное мировоззрение (просвещение и образование)

ответы на вопросы «почему», «как»

минимальный багаж знаний, **максимально** важных для каждого



базовые знания

о том, как устроен мир

методология

научного познания

рационального мышления

для каждой науки

проблематика, достижения

связи с другими науками



цивилизационная идеология (воспитание)

ответы на вопросы «ради чего», «зачем»



знание и понимание



ответы на важнейшие личные вопросы

необходимые для мотивации деятельности

система ценностей человеческой цивилизации

как прямое следствие и продолжение естественно-научной картины мира

целей и задач цивилизации

законов сохранения цивилизации

позитивных образов будущего

в чём смысл жизни

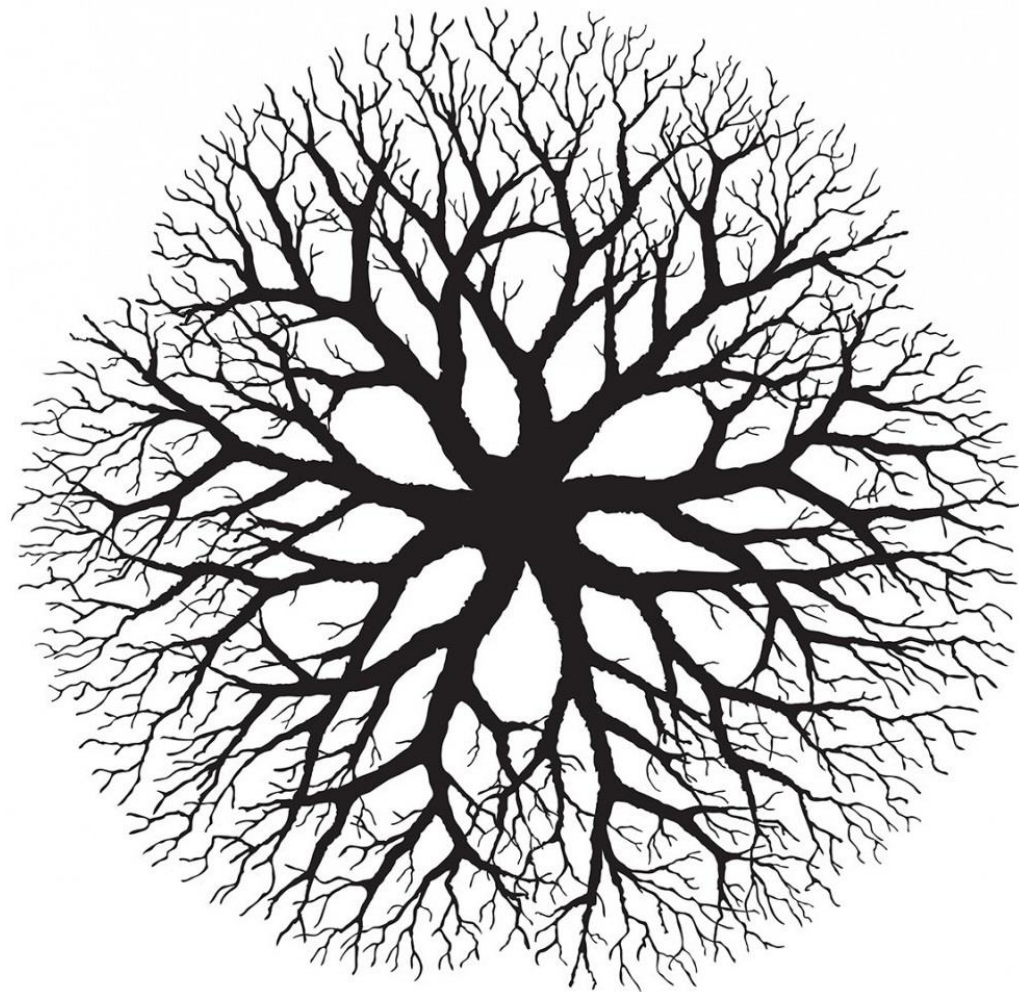
как быть счастливым

зачем нужны знания

зачем я работаю



Активация визуального аналитического мышления



1 порядка сотни карт: просмотреть, обсудить, поспорить, запомнить



2 десятки карт: построить самому, следуя 16+6 принципам



3 испытать «моменты ясности», инсайты, когда карта



индивидуальная практика и опыт

«красиво сложилась»

привела к согласию

легко и ярко запомнилась,

легла в основу деятельности

4 сделать построение карт регулярной профессиональной практикой

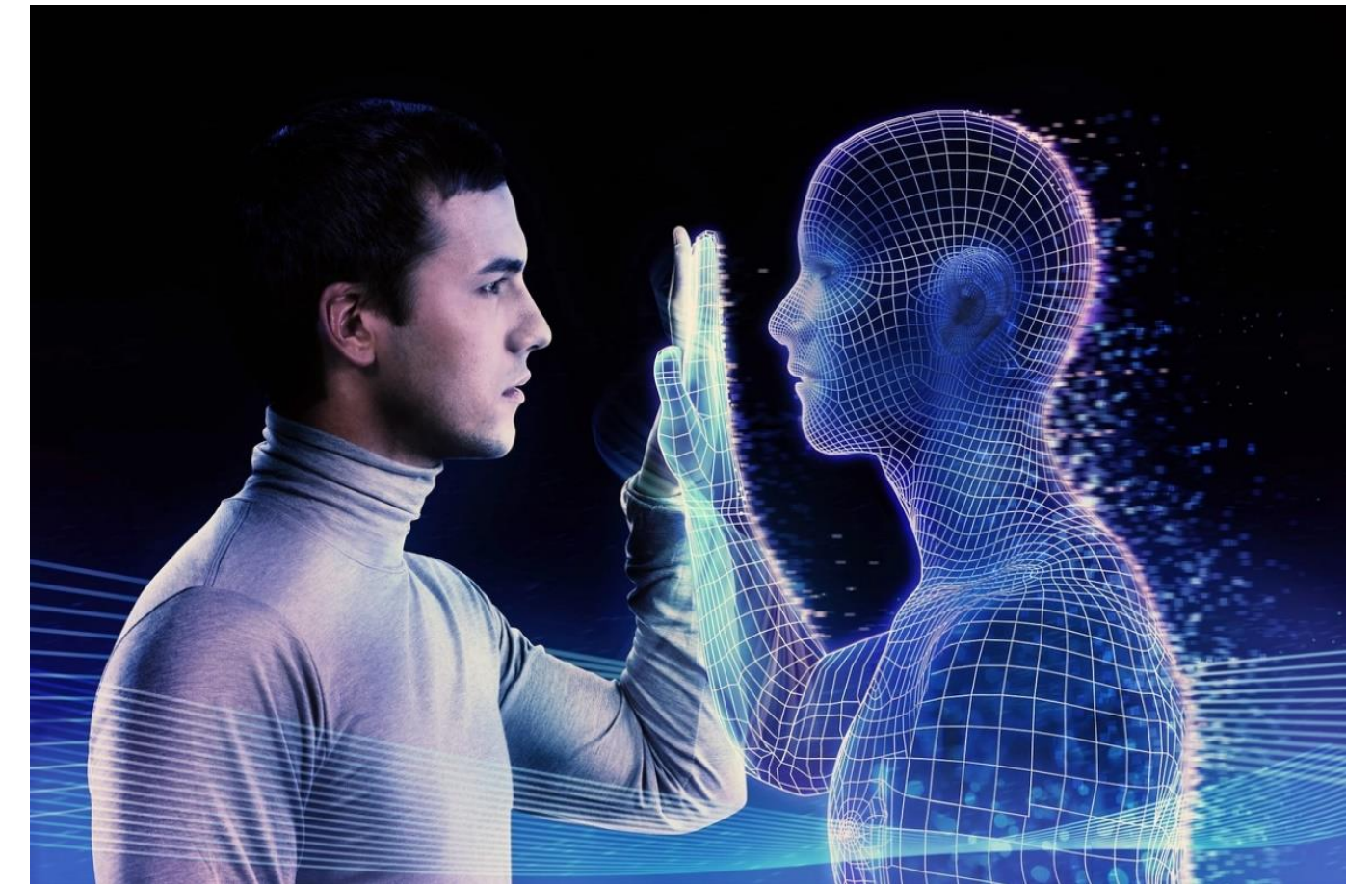


индивидуальной

коллективной

Все люди смертны

- **Чел** умирает. **Пом** становится **Аватаром**, это ценный информационный ресурс!
- **Ава** продолжает приносить пользу обществу, выполняя профессиональные и коммуникативные функции **ЧелПома**
- **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 **Человека** умирает необратимо со смертью головного мозга
- *личностный код* передаётся:

Чел → **Пом** → **Ава**



- **Ава** — это бывший **Чел**, что намного больше «фабричного ИИ», встраиваемого в машины. Он личность, 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, тревог, страхов, эгоизма
- 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)

Чел+Ава: социально-экономические изменения

- **Ава**тары станут экономическими агентами
- смешанный социум **ЧелПомов** и **Ава**таров

Риски

- полная утрата контроля, управляемости процессов развития
- новые типы социальной несправедливости и узурпации власти
- новые способы мошенничества, преступности, ведения войны

Что делать (сейчас, из нетривиального)

- разрабатывать позитивные образы наиболее вероятного будущего
- разрабатывать этику человеко-машинной цивилизации
- определять приоритетные задачи развития технологий ИИ
- ...

Некоторые вопросы этики

- **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**?
- нужно ли создавать *цивилизационное мировоззрение* или *идеологию* для человеко-машинной цивилизации людей и аватаров? (да)

Цивилизационная идеология

Иерархия цивилизационных ценностей:

- 1) биосфера Земли, уникальная в достижимой части Вселенной
- 2) человеческий вид, результат миллиардов лет эволюции
- 3) коллективный разум — образование, наука, культура
- 4) индивидуальный разум, жизнь человека
- 5) результаты труда человека

Цель человеческой цивилизации — неограниченно долгое сохранение биосферы Земли, защита её от катаклизмов ради выживания вида homo sapiens в условиях, достаточно комфортных для всех людей

Добро — всё, что этому способствует (любовь, познание, созидание, ...)

Зло — всё, что этому препятствует (ненависть, невежество, деградация, ...)



Цивилизационная
идеология
ДЗЕН-канал

<https://dzen.ru/civideology>



Какие технологии развивать? (на примере AGI)

Идём от целей и задач к технологиям, но не наоборот

- 1) каковы цели и задачи цивилизационного развития?
- 2) какие технологии необходимы и минимально достаточны?
- 3) какие задачи решаются только с помощью AGI?
- 4) генераторы текстов и картинок насколько важны для развития?
- 5) ожидаемые эффекты насколько перевешивают затраты и риски?



Любая технология — это не цель, а средство

Антропоцентричное определение ИИ — основа этики 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, творящему «по образу и подобию Своему»



Формирование позитивных образов будущего

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 цивилизационной идеологии, вывода апокалиптического сценария из исходных предположений

Идея: организовывать литературные конкурсы
в рамках проработанной фантастической гипотезы



Контакты для продолжения дискуссии



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



k.vorontsov@iai.msu.ru

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



Цивилизационная
идеология

ДЗЕН-канал

<https://dzen.ru/civideology>

