

Искусственный интеллект как фактор изменения цивилизации

Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,

рук. лаб. машинного обучения и семантического анализа Института ИИ МГУ,
зав. кафедрой математических методов прогнозирования МГУ,
зав. кафедрой машинного обучения и цифровой гуманитаристики МФТИ

k.vorontsov@iai.msu.ru

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



Три этапа развития ИИ

1. Обучение моделей по данным

— классический Machine Learning

Фрэнсис Бэкон (1620)

Карл Гаусс (1795)

Дартмутский семинар (1956) — понятие AI

Фрэнк Розенблatt (1957)

2. Обучаемая векторизация больших данных

— глубокое обучение (Deep Learning)

Джеффри Хинтон (1986, 2012)

Томаш Миколов (2013)

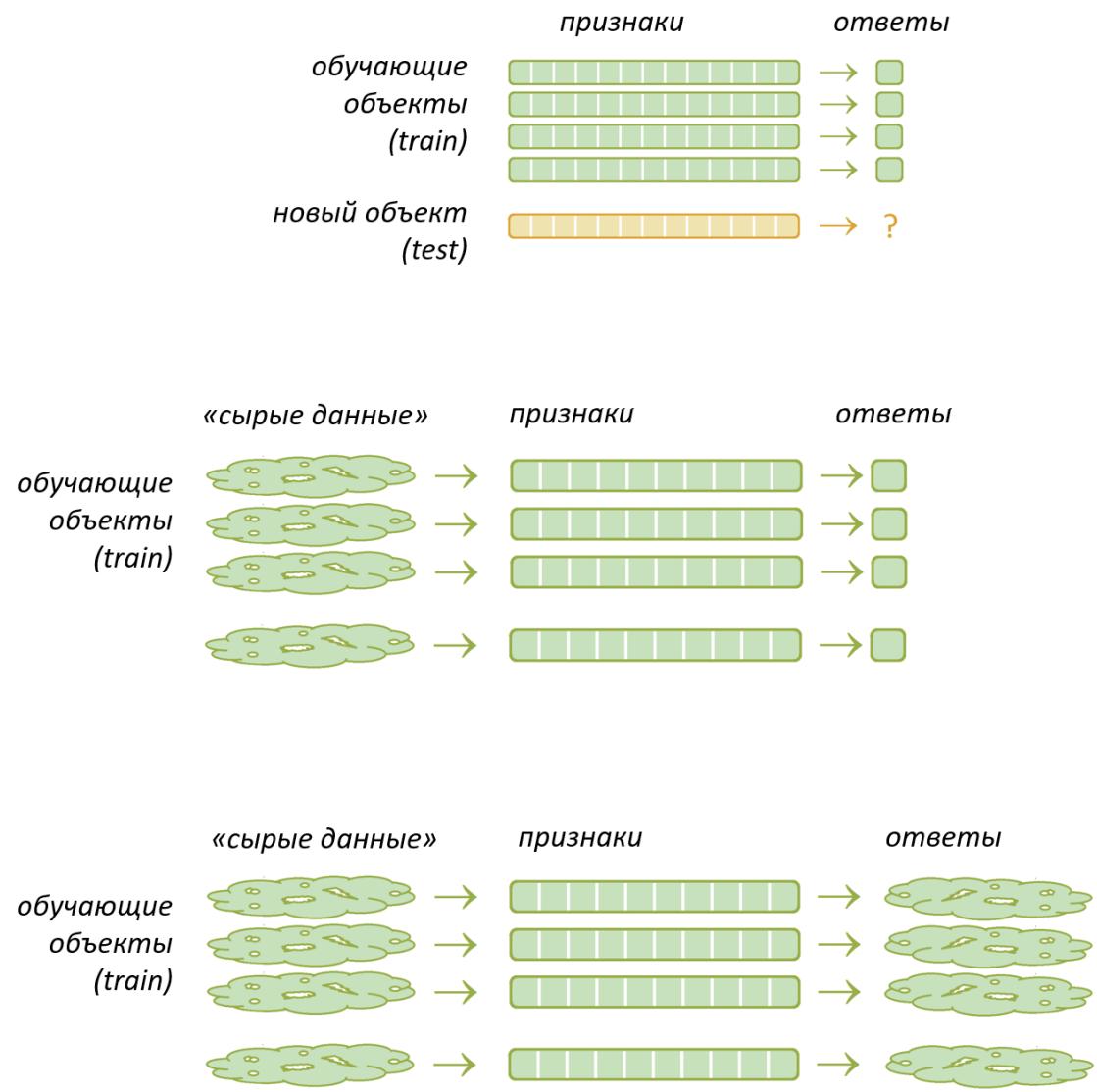
3. Генеративные модели (Generative Models)

модели внимания (2015)

трансформеры (2018)

фундаментальные модели (2021)

Midjourney (2022), GPT-4 (2023), ...



Эмерджентные способности и «проблески AGI»

Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research (27 March 2023)

Новые способности модели GPT, не закладывавшиеся при её обучении:

- объяснять свои ответы, перефразировать, переводить на другие языки
- реферировать, генерировать планы, сценарии, шаблоны
- строить аналогии, менять тональность, стиль, глубину изложения
- генерировать программный код на различных языках
- решать некоторые логические и математические задачи
- искать и исправлять собственные ошибки по подсказке

Что есть и чего нет у генеративного ИИ

Генеративная модель ИИ это

- поисковая система по коллективному знанию, накопленному человечеством
 - текстам, изображениям, музыке, видео
 - с запросами на естественном языке
 - с поисковой выдачей неточной, усреднённой, «галлюцинирующей»
 - развивающаяся, постепенно совершенствуемая разработчиками

Генеративные модели ИИ не имеют (того, что мы склонны им приписывать)

- чувств, эмоций, желаний, потребностей, мотиваций, воли
(если нам кажется, что это есть, то это наши когнитивные искажения)
- стремлений к власти, доминантности, гедонизму, лени, снятию стресса
(не имея человеческого тела, они могут говорить и показывать что угодно)
- целостной картины мира (а у нас есть?), ответственности, прав, обязанностей

Возможные сценарии будущего с ИИ

- **Персональные интеллектуальные помощники:** напоминают, рекомендуют, коммуницируют, ищут информацию, фиксируют всю жизнь своего носителя
- **Интеллектуальные агенты:** под управлением (или по сценарию) человека создают интеллектуальные продукты, выполняют бизнес-процессы
- **Автоматизация и роботизация** проектирования, производства, логистики... Люди заняты целеполаганием, управлением, контролем, сферой услуг
- **Цифровое послесмертие:** после смерти носителя помощник становится «очеловеченным ИИ», продолжая развиваться и приносить пользу
- **Человеко-машинные коллективы**
людей и аватаров решают общие задачи, аватары — не вполне люди, это новая сущность



Задачи будущего

Дополнить права человека: каждый человек имеет право

- знать, когда потребляет контент, сгенерированный ИИ
- использовать ИИ-агентов, в том числе персонального интеллектуального помощника, запоминающего его информационный поток
- на полную конфиденциальность своего информационного потока
- знать, когда коммуницирует с ИИ-агентом, а не с человеком

Создать

- **стандарты, регламенты, законы** человека-машинной цивилизации
- **способы доверенной коммуникации** между людьми и ИИ-агентами
- **единую систему знаний** доступную для людей и ИИ-агентов
- **цивилизационную идеологию**, задающую смыслы, цели, правила

Какие идеи заложены в само определение ИИ

Искусственный интеллект — интеллект, демонстрируемый машинами, в частности компьютерными системами...

— Википедия

Искусственный интеллект — комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их.

— Национальная стратегия развития искусственного интеллекта на период до 2030 года (с изменениями на 15 февраля 2024 года)

Антропоцентричное определение ИИ

Искусственный интеллект —

вычислительные технологии,
создаваемые для повышения
производительности созидательного
интеллектуального труда людей

не замена интеллекта человека

не «загадочный новый тип разума»

**не повод уподобиться Богу, чтобы
«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



k.vorontsov@iai.msu.ru

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



**Цивилизационная
идеология**
ДЗЕН-канал
<https://dzen.ru/civideology>

