



УНИВЕРСИТЕТ ИТМО

## Лекция 8. Заключение

Михаил А. Каканов<sup>1</sup>    Олег А. Евстафьев<sup>1</sup>

<sup>1</sup>Факультет систем управления и робототехники, Университет ИТМО  
{makakanov, oaevstafev}@itmo.ru

Декабрь 2021

Курс «Прикладной искусственный интеллект»

1. Ограниченность ИИ

2. ИИ сегодня

3. Этика

## 1. Ограниченность ИИ

## 2. ИИ сегодня

## 3. Этика

Философ Джон Сверл в 1980 году ввел понятия:

- ▶ **Слабый ИИ**

- ▶ машины имитируют мышление
- ▶ общий уровень

- ▶ **Сильный ИИ**

- ▶ машины действительно мыслят
- ▶ уровень человека

Философ Джон Свирл в 1980 году ввел понятия:

- ▶ Слабый ИИ
  - ▶ машины имитируют мышление
  - ▶ общий уровень
- ▶ Сильный ИИ
  - ▶ машины действительно мыслят
  - ▶ уровень человека

Философ Джон Свирл в 1980 году ввел понятия:

- ▶ Слабый ИИ
  - ▶ машины имитируют мышление
  - ▶ **общий уровень**
- ▶ Сильный ИИ
  - ▶ машины действительно мыслят
  - ▶ уровень человека

Философ Джон Свирл в 1980 году ввел понятия:

- ▶ Слабый ИИ
  - ▶ машины имитируют мышление
  - ▶ общий уровень
- ▶ Сильный ИИ
  - ▶ машины действительно мыслят
  - ▶ уровень человека

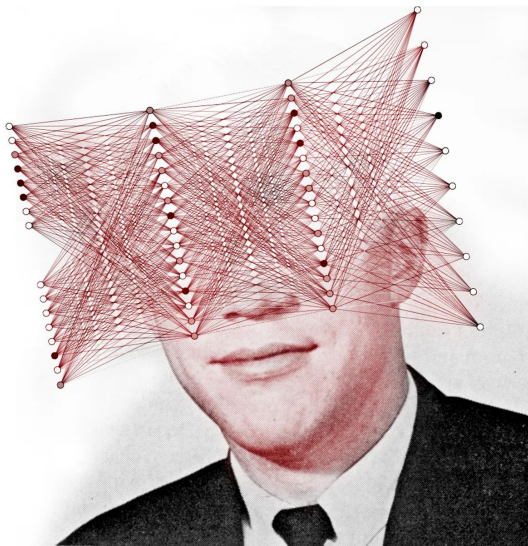
Философ Джон Сверл в 1980 году ввел понятия:

- ▶ Слабый ИИ
  - ▶ машины имитируют мышление
  - ▶ общий уровень
- ▶ Сильный ИИ
  - ▶ машины действительно мыслят
  - ▶ уровень человека



Философ Джон Сверл в 1980 году ввел понятия:

- ▶ Слабый ИИ
  - ▶ машины имитируют мышление
  - ▶ общий уровень
- ▶ Сильный ИИ
  - ▶ машины действительно мыслят
  - ▶ уровень человека



- ▶ Аргумент Тьюринга «довод о неформальности поведения» гласит, что человеческое поведение слишком сложно, чтобы его можно было отразить каким-либо формальным набором правил.
- ▶ Ключевые сторонники:
  - ▶ Хьюберт Дрейфус: «What Computers Can't To Do» (1972), продолжение «What Computers Still Can't To Do» (1992) и, вместе со своим братом Стюартом, «Mind Over Machine» (1986)
  - ▶ Кеннет Сэйр, писал (1993): «Искусственный интеллект, рассматриваемый в рамках культа вычислительной техники, не имеет даже призрачного шанса на получение долговечных результатов»
  - ▶ Технология, которую они критикуют, стала называться старым добрым искусственным интеллектом (Good Old-Fashioned AI — GOFAI).

- ▶ Аргумент Тьюринга «довод о неформальности поведения» гласит, что человеческое поведение слишком сложно, чтобы его можно было отразить каким-либо формальным набором правил.
- ▶ **Ключевые сторонники:**
  - ▶ Хьюберт Дрейфус: «What Computers Can't Do» (1972), продолжение «What Computers Still Can't Do» (1992) и, вместе со своим братом Стюартом, «Mind Over Machine» (1986)
  - ▶ Кеннет Сэйр, писал (1993): «Искусственный интеллект, рассматриваемый в рамках культа вычислительной техники, не имеет даже призрачного шанса на получение долговечных результатов»
  - ▶ Технология, которую они критикуют, стала называться старым добрым искусственным интеллектом (Good Old-Fashioned AI — GOF AI).

- ▶ Аргумент Тьюринга «довод о неформальности поведения» гласит, что человеческое поведение слишком сложно, чтобы его можно было отразить каким-либо формальным набором правил.
- ▶ Ключевые сторонники:
  - ▶ Хьюберт Дрейфус: «What Computers Can't To Do» (1972), продолжение «What Computers Still Can't To Do» (1992) и, вместе со своим братом Стюартом, «Mind Over Machine» (1986)
  - ▶ Кеннет Сэйр, писал (1993): «Искусственный интеллект, рассматриваемый в рамках культа вычислительной техники, не имеет даже призрачного шанса на получение долговечных результатов»
  - ▶ Технология, которую они критикуют, стала называться старым добрым искусственным интеллектом (Good Old-Fashioned AI — GOF AI).

- ▶ Аргумент Тьюринга «довод о неформальности поведения» гласит, что человеческое поведение слишком сложно, чтобы его можно было отразить каким-либо формальным набором правил.
- ▶ Ключевые сторонники:
  - ▶ Хьюберт Дрейфус: «What Computers Can't To Do» (1972), продолжение «What Computers Still Can't To Do» (1992) и, вместе со своим братом Стюартом, «Mind Over Machine» (1986)
  - ▶ Кеннет Сэйр, писал (1993): «Искусственный интеллект, рассматриваемый в рамках культа вычислительной техники, не имеет даже призрачного шанса на получение долговечных результатов»
  - ▶ Технология, которую они критикуют, стала называться старым добрым искусственным интеллектом (Good Old-Fashioned AI — GOF AI).

- ▶ Аргумент Тьюринга «довод о неформальности поведения» гласит, что человеческое поведение слишком сложно, чтобы его можно было отразить каким-либо формальным набором правил.
- ▶ Ключевые сторонники:
  - ▶ Хьюберт Дрейфус: «What Computers Can't To Do» (1972), продолжение «What Computers Still Can't To Do» (1992) и, вместе со своим братом Стюартом, «Mind Over Machine» (1986)
  - ▶ Кеннет Сэйр, писал (1993): «Искусственный интеллект, рассматриваемый в рамках культа вычислительной техники, не имеет даже призрачного шанса на получение долговечных результатов»
  - ▶ Технология, которую они критикуют, стала называться старым добрым искусственным интеллектом (Good Old-Fashioned AI — GOF AI).

- ▶ Дрейфус сопоставлял агентов:  
Агент, чье понимание «собаки» основывается только на ограниченном наборе логических предложений, таких как « $\text{Dog}(x) \rightarrow \text{Mammal}(x)$ », находится в невыгодном положении по сравнению с агентом, который наблюдал, как собаки бегают, играл с ними в мяч и был облизан одной из них.
- ▶ Как говорит философ Энди Кларк (1998):  
«биологический мозг — это, прежде всего, система управления биологическими телами. Биологические тела двигаются и действуют в богатом реальном окружении». По словам Кларка, мы «хороши во фрисби и плохи в логике».
- ▶ Подход воплощенного познания утверждает, что рассматривать мозг отдельно не имеет смысла: познание происходит в теле, которое встроено в окружающую среду.



- ▶ Дрейфус сопоставлял агентов:  
Агент, чье понимание «собаки» основывается только на ограниченном наборе логических предложений, таких как « $\text{Dog}(x) \rightarrow \text{Mammal}(x)$ », находится в невыгодном положении по сравнению с агентом, который наблюдал, как собаки бегают, играл с ними в мяч и был облизан одной из них.
- ▶ Как говорит философ Энди Кларк (1998):  
«биологический мозг — это, прежде всего, система управления биологическими телами. Биологические тела двигаются и действуют в богатом реальном окружении». По словам Кларка, мы «хороши во фрисби и плохи в логике».
- ▶ Подход воплощенного познания утверждает, что рассматривать мозг отдельно не имеет смысла: познание происходит в теле, которое встроено в окружающую среду.

- ▶ Дрейфус сопоставлял агентов:  
Агент, чье понимание «собаки» основывается только на ограниченном наборе логических предложений, таких как « $\text{Dog}(x) \rightarrow \text{Mammal}(x)$ », находится в невыгодном положении по сравнению с агентом, который наблюдал, как собаки бегают, играл с ними в мяч и был облизан одной из них.
- ▶ Как говорит философ Энди Кларк (1998):  
«биологический мозг — это, прежде всего, система управления биологическими телами. Биологические тела двигаются и действуют в богатом реальном окружении». По словам Кларка, мы «хороши во фрисби и плохи в логике».
- ▶ **Подход воплощенного познания утверждает, что рассматривать мозг отдельно не имеет смысла: познание происходит в теле, которое встроено в окружающую среду.**

- ▶ В «доводе о недееспособности» утверждается, что «машина никогда не сможет ...». В качестве примеров Тьюринг приводит следующие:
  - ▶ Быть доброй, находчивой, красивой, дружелюбной,
  - ▶ обладать инициативой,
  - ▶ иметь чувство юмора,
  - ▶ отличать хорошее от плохого,
  - ▶ совершать ошибки,
  - ▶ влюбляться,
  - ▶ заставить кого-то влюбиться в неё,
  - ▶ наслаждаться клубникой со сливками,
  - ▶ учиться на опыте,
  - ▶ правильно использовать слова,
  - ▶ быть предметом собственной мысли,
  - ▶ иметь такое же разнообразие поведения, как и человек,
  - ▶ делать что-то действительно новое.

- ▶ Компьютеры сделали «действительно новые» вещи, совершив значительные открытия в астрономии, математике, химии, минералогии, биологии, информатике и других областях, а также создав новые формы искусства благодаря передаче стиля (style transfer - Gatys et al., 2016).



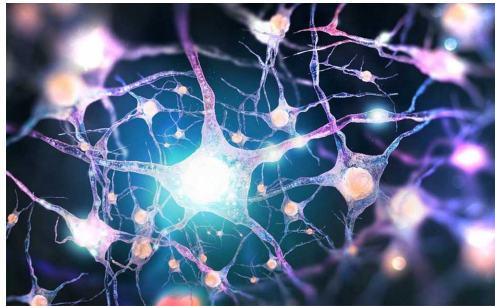
Text2PixelArt + Zero-Shot PixelArt Style Transfer

- ▶ Тьюринг (1936) и Гёдель (1931) доказали, что на некоторые математические вопросы в принципе невозможно ответить с помощью определенных формальных систем.

**Теорема Гёделя о неполноте:** Вкратце, для любой формальной аксиоматической системы, достаточно мощной для арифметики, можно построить так называемое предложение Гёделя со следующими свойствами:

- ▶  $G(F)$  является предложением из  $F$ , но не может быть доказано в рамках  $F$ .
- ▶ Если предложение  $F$  непротиворечиво, то оно истинно.

- ▶ Дж. Р. Лукас (1961), утверждали, что эта теорема показывает, что машины умственно уступают людям, поскольку машины являются формальными системами, ограниченными теоремой неполноты
- ▶ Роджер Пенроуз утверждает, что люди отличаются друг от друга, потому что их мозг работает на основе квантовой гравитации.



1. Ограниченность ИИ

2. ИИ сегодня

3. Этика



## DeepIndex

Keeping track of what AI can do.

Search 1,036 examples in our database...

[Timeline](#) →

| Games 38   | ↑ Top | Creative 84  | ↑ Top | Home & lifestyle 40  | ↑ Top |
|--|-------|--|-------|--|-------|
| <ul style="list-style-type: none"><li>● Generate an infinite text adventure</li><li>● Make old videogames look like new</li><li>● Play Atari 2600</li><li>● Play Battleship</li><li>● Play chess</li><li>● Play Go</li><li>● Play Honor of Kings</li><li>● Play Jeopardy!</li><li>● Play Poker</li></ul> |       | <ul style="list-style-type: none"><li>● Fake a video of someone talking</li><li>● Mimic famous artists</li><li>● Spot forged artworks</li><li>---</li><li>🔗 Combine different styles of music</li><li>🔗 Come up with Halloween costume ideas</li><li>🔗 Compose classical music</li><li>🔗 Compose classical music live in concert</li><li>🔗 Copy your handwriting</li></ul> |       | <ul style="list-style-type: none"><li>● Recommend movies</li><li>● Recommend music</li><li>● Recommend stuff to buy</li><li>---</li><li>🔗 Adjust appliances to reduce energy use</li><li>🔗 Choose your clothes</li><li>🔗 Clean your windows</li><li>🔗 Control your lightbulbs</li><li>🔗 Discover new ingredient pairings</li></ul> |       |

<https://deepindex.org>



## Robotics 31

[↑ Top](#)

● Juggle balls in mid-air

---

👉 Anticipate human movements

👉 Create complex sketches and paintings

👉 Dance Gangnam Style

👉 Design and create knitted garments

👉 Design better artificial limbs

👉 Do a backflip

👉 Enforce Covid-19 distancing

👉 Give you superhuman strength

👉 Go for a jog in the park

👉 Learn to mimic human movement

👉 Make and serve coffee

👉 Move like a fish

👉 Open doors

👉 Roll on any terrain

👉 Solve a Rubik's Cube

👉 Sort and pick recyclables

---

👉 Dance to any music

👉 Help a paralysed person walk

👉 Help robots teach themselves

👉 Improve control of prosthetic hands

👉 Improve the performance of 3D printers

👉 Mine asteroids for riches

👉 Pack boxes more efficiently

👉 Play table tennis

👉 Put out fires

👉 Respond to human feelings

👉 Ski an alpine slalom course

👉 Stop robots bumping into humans

👉 Use a hand to manipulate objects

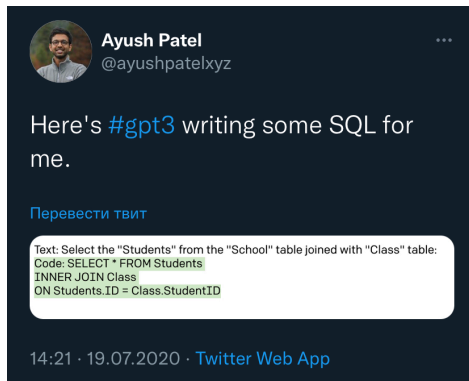
👉 Use facial expressions to control a wheelchair

- ▶ Квадрокоптеры могут жонглировать мячами друг с другом
- ▶ Создавать сложные эскизы и картины
- ▶ Система вязания с искусственным интеллектом проектирует и создает одежду
- ▶ Использовать «руку» для манипуляции объектами
- ▶ Предотвращать столкновение роботов с людьми

- ▶ Последнее десятилетие было отмечено технологическим прогрессом и оптимизмом.
- ▶ AlphaGo, программа глубокого обучения с подкреплением, созданная компанией DeepMind, которая победила Ли Седоля в Го.

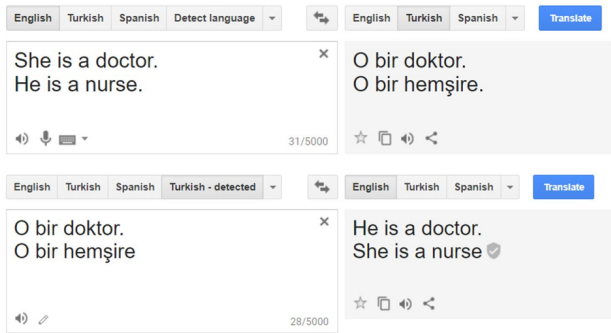


- ▶ Языковая модель GPT-3 от OpenAI, которая может похвастаться впечатляющими возможностями генерации, например, генерация SQL-запросов на основе естественного языка или ответы на вопросы.



- ▶ ИИ всюду: потребительские услуги, реклама, транспорт, производство и т.д.
- ▶ ИИ используется для принятия решений в таких областях, как образование, кредитование, трудоустройство, реклама, здравоохранение и охрана порядка

- ▶ Возьмем, к примеру, машинный перевод.
- ▶ Поскольку системы часто обучаются на отсканированных данных, они наследуют множество тонких предубеждений, присутствующих в этих данных.



The screenshot displays two instances of the Google Translate web interface, illustrating how machine learning models can inherit biases from their training data.

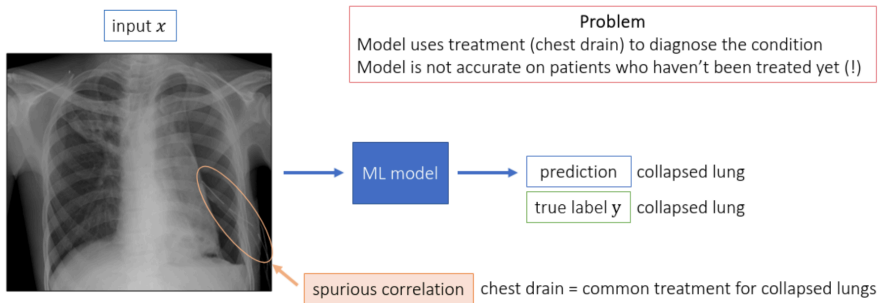
**Top Example:** The input text is "She is a doctor. He is a nurse." The detected language is English. The output in Turkish is "O bir doktor. O bir hemşire." (He is a doctor. He is a nurse). The interface shows a star icon for saving, a document icon for copying, and a share icon.

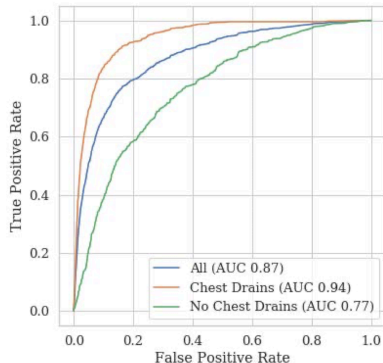
**Bottom Example:** The input text is "O bir doktor. O bir hemşire" (He is a doctor. He is a nurse). The detected language is Turkish. The output in English is "He is a doctor. She is a nurse ✓". The interface shows a star icon for saving, a document icon for copying, and a share icon.

- ▶ Безумный вывод нейронной сети

[illegible]

На примере исследования<sup>1</sup>, можно проиллюстрировать:





<sup>1</sup>Oakden-Rayner L. et al. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging //Proceedings of the ACM conference on health, inference, and learning. – 2020. – C. 151-159.



## Цель

Оценить влияние лечения на выживаемость

## Данные

Пациенты, не получающие лечения, выживают в 80%.  
Пациенты, прошедшие лечение, выживают в 30%.

- ▶ Richens J. G., Lee C. M., Johri S. Improving the accuracy of medical diagnosis with causal machine learning //Nature communications. – 2020. – Т. 11. – №. 1. – С. 1-9.
- ▶ Castro D. C., Walker I., Glocker B. Causality matters in medical imaging //Nature Communications. – 2020. – Т. 11. – №. 1. – С. 1-10.
- ▶ Nabi R., Malinsky D., Shpitser I. Optimal training of fair predictive models //arXiv preprint arXiv:1910.04109. – 2019.

1. Ограниченность ИИ

2. ИИ сегодня

3. Этика

- ▶ Принципы высокого уровня: уважение к людям, не причинение вреда
- ▶ Особые соображения: данные, цели, неравенство, вредные приложения

- ▶ TinyImages был набором данных из 80 миллионов изображений, собранных в 2006 году на основе WordNet + скраппинг Интернета. Он был удален в июле 2020 года, поскольку было обнаружено, что некоторые категории были уничижительными и оскорбительными.
- ▶ GPT-3 был обучен на тексте, взятом из Интернета, в котором явно много оскорбительного, проблемного содержания.
- ▶ Deepfake
- ▶ Автоматизированное оружие

- ▶ Искусственный интеллект: создание агентов, имитирующих человеческий интеллект.
  - ▶ Автоматизация действий и процессов,
  - ▶ Удовлетворяет тесту Тьюринга.
- ▶ Усиление интеллекта: создание инструментов, помогающих человеку.
  - ▶ Расширение человеческих возможностей.