Введение в искусственный интеллект. Современное компьютерное зрение

Семинар 10. Проблемы и вызовы современного искусственного интеллекта

Бабин Д.Н., Иванов И.Е., Петюшко А.А.

кафедра Математической Теории Интеллектуальных Систем

27 апреля 2021 г.

Что же такое искусственный интеллект?

Естественный интеллект (человек)

- Может воспринимать информацию, ее анализировать, принимать решения на основе анализа

Что же такое искусственный интеллект?

Естественный интеллект (человек)

- Может воспринимать информацию, ее анализировать, принимать решения на основе анализа

Искусственный интеллект

- (Сильный) то же самое, что и естественный, только на месте человека — компьютер

Что же такое искусственный интеллект?

Естественный интеллект (человек)

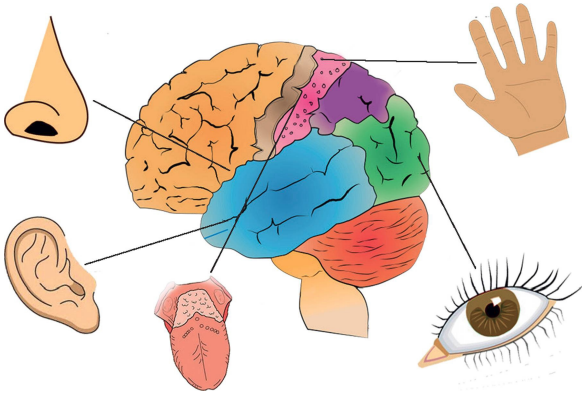
- Может воспринимать информацию, ее анализировать, принимать решения на основе анализа

Искусственный интеллект

- (Сильный) то же самое, что и естественный, только на месте человека — компьютер
- (**Слабый**) алгоритм, способный обучиться на основе массива входных данных, чтобы затем выполнять задачу вместо человека

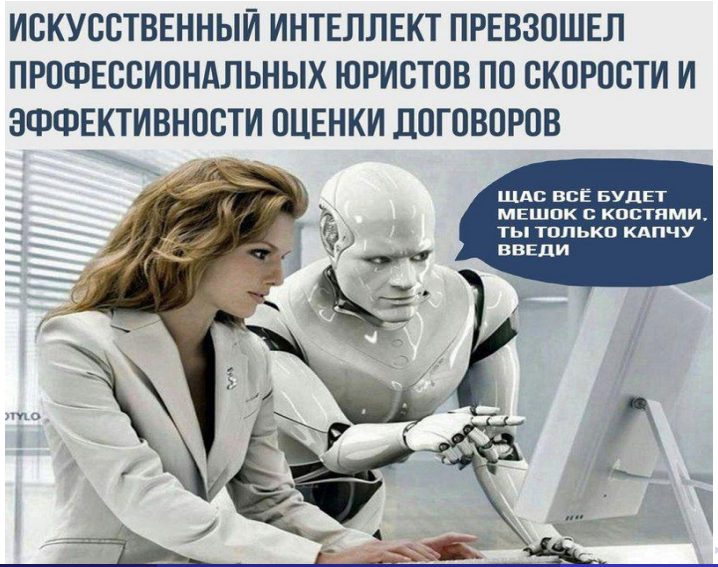
Взаимодействие со средой

- Около 90 % информации поступает через **зрение**[1]
- Около 9 % информации поступает через слух



[1] https://www.rlsnet.ru/books_book_id_2_page_40.htm

Успехи искусственного интеллекта



Успехи искусственного интеллекта

- Самоуправляемые автомобили
- Магазины без продавцов
- Голосовые помощники
- Рекомендательные системы
- Интернет вещей
- Умный дом
- Умный город

Успехи искусственного интеллекта

- Самоуправляемые автомобили
- Магазины без продавцов
- Голосовые помощники
- Рекомендательные системы
- Интернет вещей
- Умный дом
- Умный город

Вывод

Влияние искусственного интеллекта на жизнь человека огромно и оно продолжает расти

ИИ в жизни

Вы могли не замечать, но уже

- Вы смотрите видео, которое рекомендует вам youtube
- Вы передвигаетесь по маршруту, который вам построил навигатор
- В банке робот решает давать вам кредит или нет
- Штрафы за нарушение ПДД выписываются автоматически
- ИИ гугла ранжирует вам доступ к информации
- ...

ИИ в жизни

Вы могли не замечать, но уже

- Вы смотрите видео, которое рекомендует вам youtube
- Вы передвигаетесь по маршруту, который вам построил навигатор
- В банке робот решает давать вам кредит или нет
- Штрафы за нарушение ПДД выписываются автоматически
- ИИ гугла ранжирует вам доступ к информации
- ...

Вывод

Искусственный интеллект уже принимает активное участие в вашей жизни!

Этика искусственного интеллекта

Вопросы

- Что будет если ИИ начнет себя вести не совсем этично по отношению к некоторым группам населения?

Этика искусственного интеллекта

Вопросы

- Что будет если ИИ начнет себя вести не совсем этично по отношению к некоторым группам населения?
- Кто будет за это отвечать?

Этика искусственного интеллекта

Вопросы

- Что будет если ИИ начнет себя вести не совсем этично по отношению к некоторым группам населения?
- Кто будет за это отвечать?
- Как вообще определить, что такое происходит и происходит ли?

Этика искусственного интеллекта

Вопросы

- Что будет если ИИ начнет себя вести не совсем этично по отношению к некоторым группам населения?
- Кто будет за это отвечать?
- Как вообще определить, что такое происходит и происходит ли?
- Как такое предотвратить?

Пример дискриминации

Пример

Объявление о высокооплачиваемой работе чаще показываются белым мужчинам среднего возраста

Пример дискриминации

Пример

Объявление о высокооплачиваемой работе чаще показываются белым мужчинам среднего возраста

Причины

- Потому что компания скорее всего наймет белого мужчину среднего возраста

Пример дискриминации

Пример

Объявление о высокооплачиваемой работе чаще показываются белым мужчинам среднего возраста

Причины

- Потому что компания скорее всего наймет белого мужчину среднего возраста
- Есть функция потерь, и ИИ ее оптимизирует. В данном случае дискриминирование эквивалентно эффективному показу рекламы

Пример дискриминации

Пример

Объявление о высокооплачиваемой работе чаще показываются белым мужчинам среднего возраста

Причины

- Потому что компания скорее всего наймет белого мужчину среднего возраста
- Есть функция потерь, и ИИ ее оптимизирует. В данном случае дискриминирование эквивалентно эффективному показу рекламы

Возможное решение

- Если в нанимающей компании есть квоты на дискриминируемые меньшинства, то такой алгоритм выдачи перестаёт быть эффективным

Ответственность перед обществом

Вопрос

Кто ответственен за подобное поведение ИИ?

Ответственность перед обществом

Вопрос

Кто ответственен за подобное поведение ИИ?

Возможный ответ

Компания, которая производит такой продукт

Ответственность перед обществом

Вопрос

Кто ответственен за подобное поведение ИИ?

Возможный ответ

Компания, которая производит такой продукт

Возможное возражение

Компания, которая производит такой продукт, настраивается на данные, за которые она не отвечает

Ответственность перед обществом

Вопрос

Кто ответственен за подобное поведение ИИ?

Возможный ответ

Компания, которая производит такой продукт

Возможное возражение

Компания, которая производит такой продукт, настраивается на данные, за которые она не отвечает

Вопрос

Можем ли требовать от ИИ быть более этичным, чем само общество в целом?

Второй пример дискриминации

Пример

Банк не был готов дать кредит супругам вместе. Вместо этого предлагал заключить брачный договор. Хотя кредитный рейтинг жены был выше, чем у мужа. На 100% нельзя утверждать, что тут был случай дискриминации, но банк отказался прокомментировать своё решение.

Второй пример дискриминации

Пример

Банк не был готов дать кредит супругам вместе. Вместо этого предлагал заключить брачный договор. Хотя кредитный рейтинг жены был выше, чем у мужа. На 100% нельзя утверждать, что тут был случай дискриминации, но банк отказался прокомментировать своё решение.

Непрозрачность

В данном случае непрозрачность системы принятия решений позволяет ИИ дискриминировать.

Интерпретируемость ИИ

Проблема

Во многих случаях ИИ работает как черный ящик. В некоторых областях такое поведение недопустимо

Интерпретируемость ИИ

Проблема

Во многих случаях ИИ работает как черный ящик. В некоторых областях такое поведение недопустимо

Пример

Медицина

Третий пример дискриминации

Распознавание лиц

Известный факт, что на африканцах распознавание лиц работает хуже, чем на европейцах

Причины

Датасеты для европейцев более представительные.

Самоуправляемые автомобили

- Этот тип ИИ вызывает наиболее оживленную дискуссию

Самоуправляемые автомобили

- Этот тип ИИ вызывает наиболее оживленную дискуссию
- Самоуправляемые автомобили уже на дорогах Москвы

Самоуправляемые автомобили

- Этот тип ИИ вызывает наиболее оживленную дискуссию
- Самоуправляемые автомобили уже на дорогах Москвы
- Известны случаи, когда люди гибли по вине самоуправляемых автомобилей

Самоуправляемые автомобили

- Этот тип ИИ вызывает наиболее оживленную дискуссию
- Самоуправляемые автомобили уже на дорогах Москвы
- Известны случаи, когда люди гибли по вине самоуправляемых автомобилей
- Известны случаи, когда люди были спасены ими

Самоуправляемые автомобили

- Этот тип ИИ вызывает наиболее оживленную дискуссию
- Самоуправляемые автомобили уже на дорогах Москвы
- Известны случаи, когда люди гибли по вине самоуправляемых автомобилей
- Известны случаи, когда люди были спасены ими
- Перед ИИ стоит множество этических вопросов: как вести в той или иной ситуации

Самоуправляемые автомобили

- Этот тип ИИ вызывает наиболее оживленную дискуссию
- Самоуправляемые автомобили уже на дорогах Москвы
- Известны случаи, когда люди гибли по вине самоуправляемых автомобилей
- Известны случаи, когда люди были спасены ими
- Перед ИИ стоит множество этических вопросов: как вести в той или иной ситуации

Проблема

Рынок труда изменится при всеобщем внедрении этой технологии.

Проблема

Добавление малозаметного шума к изображению ставит нейронную сеть в тупик

Атаки на ИИ в реальной жизни

Проблема

Первые три знака распознаются как «Ограничение скорости 45», а последний — как знак «STOP»

Атаки на ИИ в реальной жизни

Проблема

Специальные элементы одежды (очки, шапки и т.д.) позволяют обмануть систему распознавания лиц

Данные vs Частная жизнь

Некторые факты

- Сотовые операторы продают данные о местоположении пользователей

Данные vs Частная жизнь

Некторые факты

- Сотовые операторы продают данные о местоположении пользователей
- Говорят, что Амазон по вводу запроса может определять беременность

Данные vs Частная жизнь

Некторые факты

- Сотовые операторы продают данные о местоположении пользователей
- Говорят, что Амазон по вводу запроса может определять беременность
- Ваши поисковые запросы влияют на ваши рекомендации

Данные vs Частная жизнь

Некторые факты

- Сотовые операторы продают данные о местоположении пользователей
- Говорят, что Амазон по вводу запроса может определять беременность
- Ваши поисковые запросы влияют на ваши рекомендации

Вопрос

Кому должны пренадлежать данные, производимые людьми? Какие данные являются приватными, а какие нет?

Безопаность vs Частная жизнь

Обеспечение безопасности

Для обеспечения безопасности государство устанавливает камеры видеонаблюдения

Безопаность vs Частная жизнь

Обеспечение безопасности

Для обеспечения безопасности государство устанавливает камеры видеонаблюдения

Вопрос

Где границы того, что должно быть доступно государству, а что нет?

Пример принципов этики ИИ

- Secure AI. Контролируемость и управляемость систем ИИ
- Explainable AI. Прозрачность и предсказуемость функционирования
- Reliable AI. Стабильность и надежность систем ИИ
- Responsible AI. Ответственное применение ИИ
- Fair AI. Непредвзятый ИИ

Заключение

- Жизнь современного человека невозможно предствавить без ИИ
- Сейчас нет общего понимания и согласия, каким должен быть этичный ИИ
- Но уже многие компании задумываются об этике ИИ
- Исследователи и разработчики должны больше обращать внимание на проблемы этики ИИ

Спасибо за внимание!