Технологии разработки интеллектуальных аналитических систем

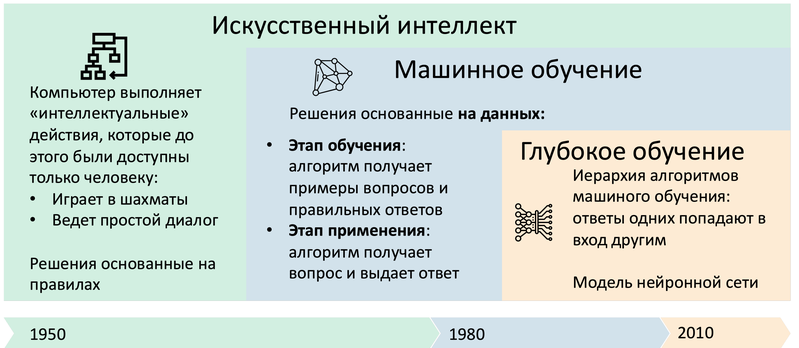
Денис Турдаков
ИСП РАН

Опыт ИСП РАН

- Инструменты поддержки жизненного цикла безопасной разработки ПО
- Инструменты анализа данных
  - Анализ больших данных
  - Анализ соц. сетей
  - Анализ естественного языка
  - Анализ изображений
  - Анализ временных рядов
  - Физическое моделирование
  - …
- Инструменты поддержки ЖЦ безопасной разработки интеллектуальных систем
- Индустриальные исследования и внедрение результатов
- Базовые кафедры ВМК МГУ, ФПМИ МФТИ, ФКН ВШЭ
- Три лаборатории за пределами ИСП РАН

В 2021 году на базе ИСП РАН создан Центр доверенного ИИ – один из 6 центров, поддержанных грантами в рамках федерального проекта «Искусственный интеллект»

Технологии искусственного интеллекта



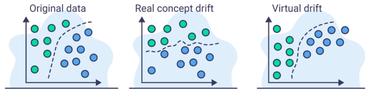
Переход от построения модели объекта автоматизации к решению проблемы по аналогии

Вызовы и требования к интеллектуальным системам

Вызовы использования ИИ	***Возможные решения***
Естественные ошибки ИИ	Разработка **автоматизированных** систем (для области, где ошибка критична)
Постоянные **революционные изменения** в области ИИ Конкурентоспособность системы определяется новизной используемых решений	Архитектура решений должна позволять быстро внедрять новые технологии и модели (time-to-market)
Необходимость учитывать **дрейф данных*** при построении отчуждаемых решений и при ограничениях доступа к данным заказчика * Мир постоянно меняется, поэтому модели машинного обучения начинают устаревать в момент окончания разработки	Инфраструктура поддержки полного жизненного цикла машинного обучения становится частью интеллектуальных систем
Возникновение принципиально **новых угроз**, связанных с использованием машинного обучения	В инфраструктуру добавляются методики и инструменты безопасной разработки

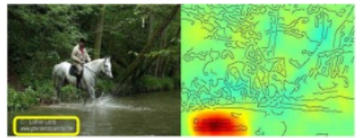
Проблемы доверия к системам с ИИ

Дрейф данных



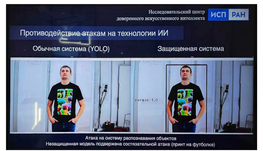
Предвзятость

Неитерпретируемость



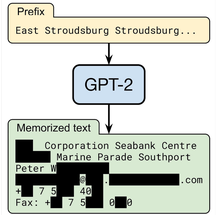
Проблемы доверия к системам с ИИ

Состязательные атаки



Встраивание закладок

Извлечение конфиденциальных данных



Платформа «Talisman»

Talisman – платформа для построения интеллектуальных информационно-аналитических систем

Извлечение информации

- Построение базы знаний на основе анализа неструктурированных данных (текст, изображения, видео, аудио, графы, web-страницы)
- Анализ более 100 языков
- Сбор данных из Интернета и корпоративных хранилищ

Аналитические инструменты

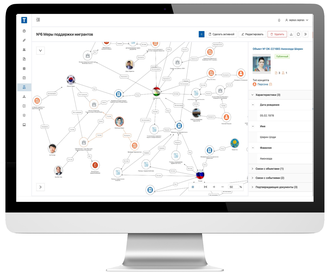
- Быстрый поиск и фильтрация информации с использованием накапливаемых знаний
- Построение дашбордов на основе информации из неструктурированных данных
- Проверка информации (fact checking)

Внедрения

- 300+ млн. руб., в т.ч. от продажи лицензий
- Ведение корпоративных баз знаний
- Конкурентная и технологическая разведка, оценка контрагентов
- Анализ данных в области международных отношений
- Системы мониторинга информации в сети Интернет

>50 моделей **машинного обучения**
анализ текста, изображений, видео, графов, таблиц

В реестре российского ПО



Машинное обучение в «Talisman»

Настраиваемая предметная область

- Снижение числа **естественных ошибок** моделей извлечения информации
- Интеграция с технологиями semantic web и пополнение базы на основе Linked Open Data

TALISMAN MAGIC

Machine Learning Automated Governance and Integration Controller

- Оператор может верифицировать факты, попадающие в базу знаний
- Естественная разметка датасетов (human-in-the-loop)
- Система производит тестирование **обучение и дообучение моделей** на этих данных
- Решает проблему **дрейфа данных**

Модульная архитектура и универсальная модель данных

- Микросервисы и эластичное масштабирование для обработки больших данных
- Позволяет **быстро внедрять** SOTA модели извлечения информации

Решаемые адачи

- классификация документов
- извлечение именованных сущностей и отношений
- обработка таблиц
- классификация изображений
- оптическое распознавание символов
- анализ эмоциональной окраски
- распознавание лиц
- обучение ранжированию
- анализ графа знаний и др.

- https://github.com/ispras/dedoc
- Сервис для перевода документов в универсальный формат
- Позволяет учитывать структуру документа при обработке
- Обрабатывает docx, odt, json, pdf, html, jpg, png, zip, rar…
- https://habr.com/ru/companies/isp_ras/articles/779390/

Платформа для анализа и разработки доверенных интеллектуальных систем

- Основной программный продукт Исследовательского центра доверенного искусственного интеллекта ИСП РАН
- Апробация индустриальными партнерами Центра

Интеграция с Talisman

- Доверенная версия платформы «Talisman»
- Анализ используемых датасетов
- Анализ моделей машинного обучения

Инструменты разработчика

- Методы защиты моделей от атак на этапе эксплуатации
- Методы обнаружения дрейфа данных
- Инструменты оценки устойчивости обученных моделей к атакам
- Методы объяснения моделей
- Инструменты для повышения доверия к предобученным моделям
- Методы выявления предвзятости моделей
- Инструменты проверки наличия аномалий в наборах данных
- Доверенные версии TensorFlow и PyTorch

Спасибо!

ИСП РАН

ЦИФРА СТРОЙ

XV Академические чтения, посвященные памяти академика РААСН Осипова Г.Л.

Научно-практическая конференция «Перспективы использования искусственного интеллекта в градостроительной деятельности», Москва, 2 – 3 июля 2024 г.

Модераторы:

Валерия Мозганова, Радиостанция Business FM, руководитель отдела «Недвижимость»

Евгений Карант, НИИСФ РААСН, ведущий инженер

Полный список докладов доступен на сайте ЦифраСтрой по ссылке

https://cifrastroy.ru/news/buduschee-iskusstvennogo-intellekta-v-gradostroitelstve