



Правосудие по алгоритму: искусственный интеллект в работе полиции и уголовном правосудии

Борис Цилевич

Санкт-Петербург,

4.11.2019

Информационные технологии - ИТ

- Обработка данных
- Алгоритмы: детерминированные и недетерминированные (*список покупок*)
- «Чёрный ящик»
- Права человека и верховенство закона?
- **Проблема ответственности**

Искусственный интеллект

- **«Сильный»**: превосходит человека во всех областях
- **«Слабый»**: способен выполнять отдельные задачи с определённой степенью автономии
- Нет универсального и общепринятого определения
- Моделирование человеческого интеллекта
- Ключевой элемент: результат и поведение алгоритма не определяется полностью программистом, алгоритм «эволюционирует», меняется и развивается на основе накопленного «опыта»
- Базы данных – базы знаний. Алгоритм «учится»

Правовое регулирование

- Эффект «чёрного ящика»: не только из-за ограниченных «мощностей» человека, но и из-за принципиальной непредсказуемости «эволюционирующего» алгоритма
- Автономия – ответственность?
- Этические кодексы и стандарты
- Управление рисками
- Доверие общества
- Уровень регулирования: национальный vs международный. *“Ethics shopping”*
- **Совет Европы:** разработка стандартов в области прав человека, демократии и верховенства права

CEPEJ 2018

EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ)

European Ethical Charter
on the Use of Artificial Intelligence in
Judicial Systems and their environment



Adopted at the 31th plenary meeting
of the CEPEJ on 3-4 December 2018

СЕРЕЈ 2018

- **Основные принципы Хартии:**
- **Уважение фундаментальных прав**
 - разработка и внедрение искусственного интеллекта не должна нарушать фундаментальные права человека
- **Принцип недискриминации**
 - предотвращение появления или усиления дискриминации в отношении отдельных людей и групп
- **Принцип качества и безопасности**
 - обработка судебных решений и данных должна быть технически защищённой, проводиться на основании проверенных источников и надёжных и адекватных моделей
- **Принцип открытости, беспристрастности и честности**
 - методы обработки данных должны быть доступны и понятны
- **Принцип «под контролем пользователя»**
 - пользователям должно быть обеспечено право выбора и доступ к необходимой информации

Council of Europe Commissioner for Human Rights 2019

Unboxing Artificial Intelligence: 10 steps to protect Human Rights



Council of Europe Commissioner for Human Rights 2019

- Оценка влияния на права человека
- Консультации с обществом
- Обязательство государств содействовать соблюдению прав человека в частном секторе
- Доступность информации и прозрачность
- Независимый контроль
- Недискриминация и равенство
- Защита личных данных и частной жизни
- Свобода слова, свобода собраний и ассоциаций, право на работу
- Механизмы устранения нарушений
- Содействие «ИИ-грамотности»

Другие органы Совета Европы

- GRECO
 - борьба с отмыванием денег и финансированием терроризма
- DH-BIO
 - биоэтика
- «Группа Помпиду»
 - борьба со злоупотреблением наркотиками и их незаконным оборотом
- Венецианская комиссия
 - цифровые технологии и выборы
- ECRI
- Гендерное равенство, молодёжная политика, образование, культура...

Парламентская Ассамблея

- *Recommendation 2069 (2015), Resolution 2051 (2015):* Drones and targeted killings: the need to uphold human rights and international law
- *Recommendation 2102 (2017):* Technological convergence, artificial intelligence and human rights

Парламентская Ассамблея

- **Committee on legal affairs and human rights**
 - Justice by algorithm – the role of artificial intelligence in policing and criminal justice systems
 - The human being-machine connection: new rights or new threats to fundamental freedoms?
 - Legal aspects of "autonomous" vehicles
 - Sub-committee on AI and HR
- **Committee on political affairs and democracy**
 - Need for democratic governance of artificial intelligence
- **Committee on social affairs, health, sustainable development**
 - Artificial intelligence and labour markets: friend or foe?
- **Committee on equality and non-discrimination**
 - Preventing discrimination caused by the use of artificial intelligence
- **Committee on culture, science, education and media**
 - General Rapporteur on science and technology assessment

ИИ в правосудии: PredPol

- Прогнозирование правонарушений и оптимизация распределения ресурсов
- Разработка UCLA и Los Angeles Police Department
- Определение места и времени вероятных правонарушений и планирование маршрутов патрулирования
- Для «обучения» алгоритмов используются данные за предыдущие 2-5 лет
- Используются **только** данные о
 - виде правонарушения
 - месте совершения
 - дате и времени

ИИ в правосудии: PredPol

- Использовалась полицией Кента (UK) с 12.2012 до 03.2018.
- Снижение уличной преступности на 6%
- Стоимость: 100,000 £ в год
- **Критика:** усиление и увековечивание исторически сложившейся предвзятости - *tech-washing*
- Ответ на критику: проект NDAS (National data Analytics Solution)
 - West Midlands Police, London Metropolitan Police, Greater Manchester Police...
 - Оценка риска правонарушений с использованием огнестрельного или холодного оружия
 - 1,400 индикаторов для прогнозирования преступлений
 - Участие UK Data Protection authority, независимая оценка и контроль

ИИ в правосудии: HART

- Прогнозирование повторных правонарушений и предотвращение рецидива
- Разработка Durham Police и Cambridge University
- Данные о 100,000 лиц, арестованных за последние 5 лет
- Оценка уровня риска совершения новых правонарушений
- **Критика** - аналогична PredPol: предикторы провоцируют увековечение существующих паттернов (в частности, адреса и почтовые индексы)
- Ответ на критику со стороны Durham Police: HART используется лишь в качестве консультанта, все индивидуальные решения принимают профессионально подготовленные офицеры полиции

ИИ в правосудии: COMPAS

- Прогнозирование повторных правонарушений и содействие в принятии решений о мерах пресечения, вынесении приговоров и принятии решений о досрочном освобождении
- Разработка компании Equivant, используется в нескольких штатах США в качестве «инструмента оценки рисков и потребностей»
- **Критика:** исследовательский сайт *Pro Publica*:
 - «Алгоритм даёт несколько большую точность, чем подбрасывание монеты»
 - Процент ошибок в оценке уровней риска для белых и чёрных примерно одинаков, но алгоритм гораздо чаще даёт ошибочную оценку высокого риска для чернокожих и ошибочную оценку низкого риска – для белых
- Впрочем, методики и выводы *Pro Publica* также подверглись обоснованной критике
 - «Предвзятость не в инструменте, а в системе»

В заключение

- ИИ разрабатывается с середины 20 века, но лишь в последние годы резко возросло его практическое применение в различных областях, в том числе в правосудии
- Очевидно, что использование ИИ может очень существенно повлиять на соблюдение прав человека, в особенности в уголовном судопроизводстве
- В условиях глобализации и принципиально транснационального характера разработки ИИ любое регулирование на национальном уровне будет недостаточно эффективным
- Таким образом, необходимо тесное международное сотрудничество в разработке стандартов применения ИИ, гарантирующих учёт и соблюдение прав человека, и механизмов их практического внедрения.
- Рамочная конвенция?