

НОЦ «ФНС России и МГТУ им. Н.Э. Баумана»
НОЦ «Технологии ИИ» МГТУ им. Н.Э. Баумана

Интерпретируемые методы искусственного интеллекта и их роль в государственном управлении и финансовом секторе



ФНС
ЛАБ

НОЦ ФНС РОССИИ
И МГТУ ИМ. Н.Э. БАУМАНА



Тынченко
Вадим
Сергеевич

Главный научный сотрудник,
д.т.н., доцент

29 мая 2024 года

Сферы применения технологий искусственного интеллекта



ФИНАНСЫ

ПРЕДОТВРАЩЕНИЕ МОШЕННИЧЕСТВА
ОБСЛУЖИВАНИЕ КЛИЕНТОВ
КОНТРОЛЬ ПРАВОМЕРНОСТИ ДЕЙСТВИЙ
БАНКА



ОБРАЗОВАНИЕ

БЕСПРИСТРАСТНОСТЬ ОЦЕНИВАНИЯ
ИНДИВИДУАЛЬНЫЕ ПРОГРАММЫ
РЕАЛИЗАЦИЯ ПОТЕНЦИАЛА УЧЕНИКА



ОРГАНИЗАЦИЯ СБЫТА

СОЗДАНИЕ КОМФОРТНЫХ УСЛОВИЙ
ПРЕДОТВРАЩЕНИЕ АВАРИИ ЖКХ
ПОМОЩЬ ЛЮДЯМ С ОГРАНИЧЕННЫМИ
ВОЗМОЖНОСТЯМИ



СТРУКТУРА БЕЗОПАСНОСТИ

ВЫПОЛНЕНИЕ ОПАСНЫХ ДЛЯ ЖИЗНИ
ЧЕЛОВЕКА ФУНКЦИЙ
ПРЕДУПРЕЖДЕНИЕ ПРЕСТУПЛЕНИЙ
БЕСПРИСТРАСТНОСТЬ НАКАЗАНИЯ

МЕДИЦИНА

РЕГУЛЯРНАЯ ДИАГНОСТИКА ЗДОРОВЬЯ
РАЗРАБОТКА ЛЕКАРСТВ
ТЕЛЕМЕДИЦИНА



СЕЛЬСКОЕ ХОЗЯЙСТВО

АВТОМАТИЗАЦИЯ СБОРА УРОЖАЯ
БОРЬБА С СОРНЯКАМИ И НАСЕКОМЫМИ
ПОДДЕРЖКА "ТЕПЛИЧНЫХ" УСЛОВИЙ



ПРОМЫШЛЕННОСТЬ

АВТОМАТИЗАЦИЯ
СНИЖЕНИЕ ИЗДЕРЖЕК
ПОВЫШЕНИЕ КАЧЕСТВА ПРОДУКТА



ТРАНСПОРТ

БЕЗОПАСНОСТЬ НА ДОРОГАХ
РАСПИСАНИЕ ТРАНСПОРТА ОНЛАЙН
РАЗГРУЗКА ДОРОЖНЫХ «ПРОБОК»



Национальная стратегия развития искусственного интеллекта

Россия должна стать одним из международных лидеров в развитии ИИ для

- 1 Роста благосостояния и качества жизни населения
- 2 Стимулирования экономического развития
- 3 Обеспечения национальной безопасности и охраны правопорядка

Для реализации поставленной цели необходимо способствовать внедрению ИИ через взаимодействие бизнеса, государства, науки и общества

Доля крупных и средних компаний, органов государственной власти и подведомственных организаций, использующих ИИ:

- К 2024 году – 10%
- К 2030 году – 20%

ОСНОВНЫЕ ДОКУМЕНТЫ ВЫСОКОТЕХНОЛОГИЧНОЙ ОТРАСЛИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»

Указ Президента Российской Федерации
от 10 октября 2019 №490
«О развитии искусственного интеллекта
в Российской Федерации»



ФЗ от 24.04.2020 г. № 123-ФЗ
О специальном регулировании в целях
создания необходимых условий для
разработки и внедрения технологий ИИ в
Москве



Тренды ИИ



ГЕНЕРАТИВНЫЙ ИИ



Midjourney

ОБЪЯСНИМЫЙ ИИ

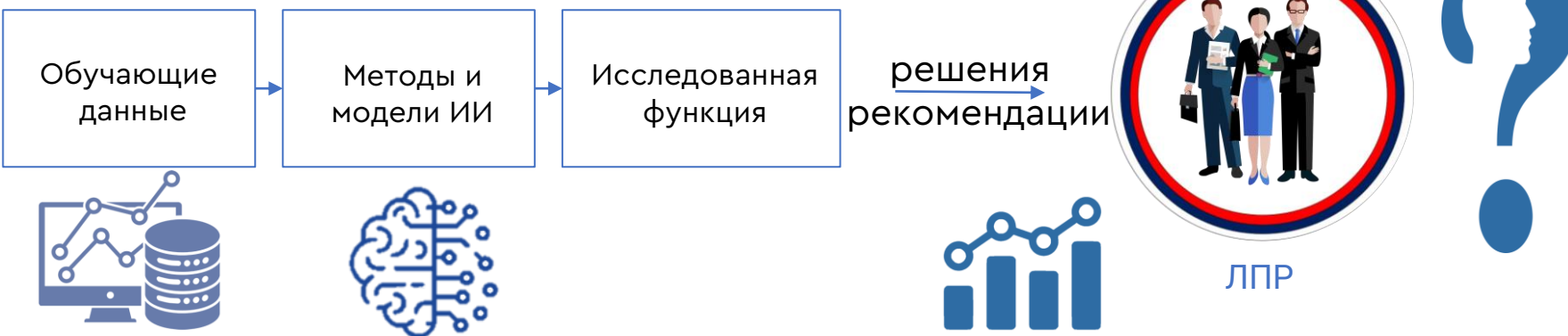
ВИРТУАЛЬНЫЕ ПОМОЩНИКИ



ПЕРИФЕРИЙНЫЙ ИИ (ИНТЕРНЕТ ВЕЩЕЙ)

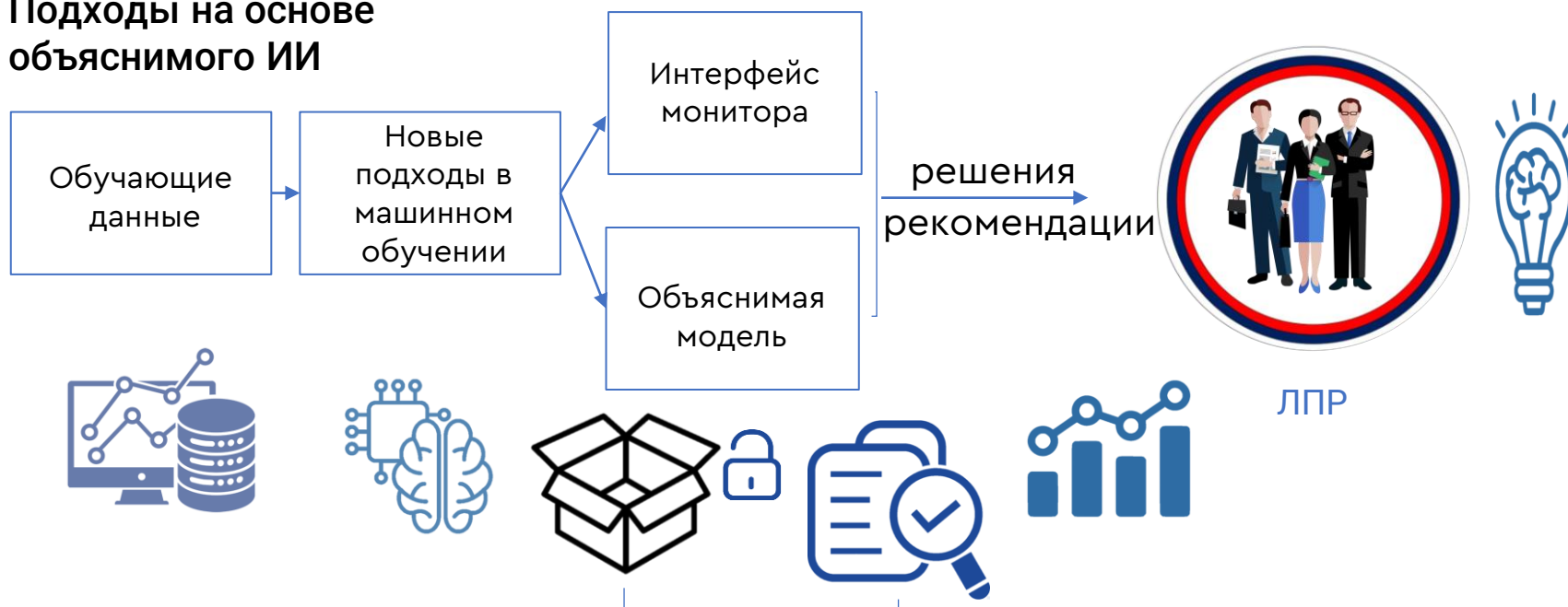
Роль и место интерпретируемого ИИ в государственном управлении и финансовом секторе

Применяемые сегодня подходы ИИ



Почему ИИ так поступил?
Почему именно это решение?
Когда ИИ добивается результата?
Когда ИИ терпите неудачу?
Когда можно доверять решению ИИ?
Как исправить ошибку на основе «черного ящика»?

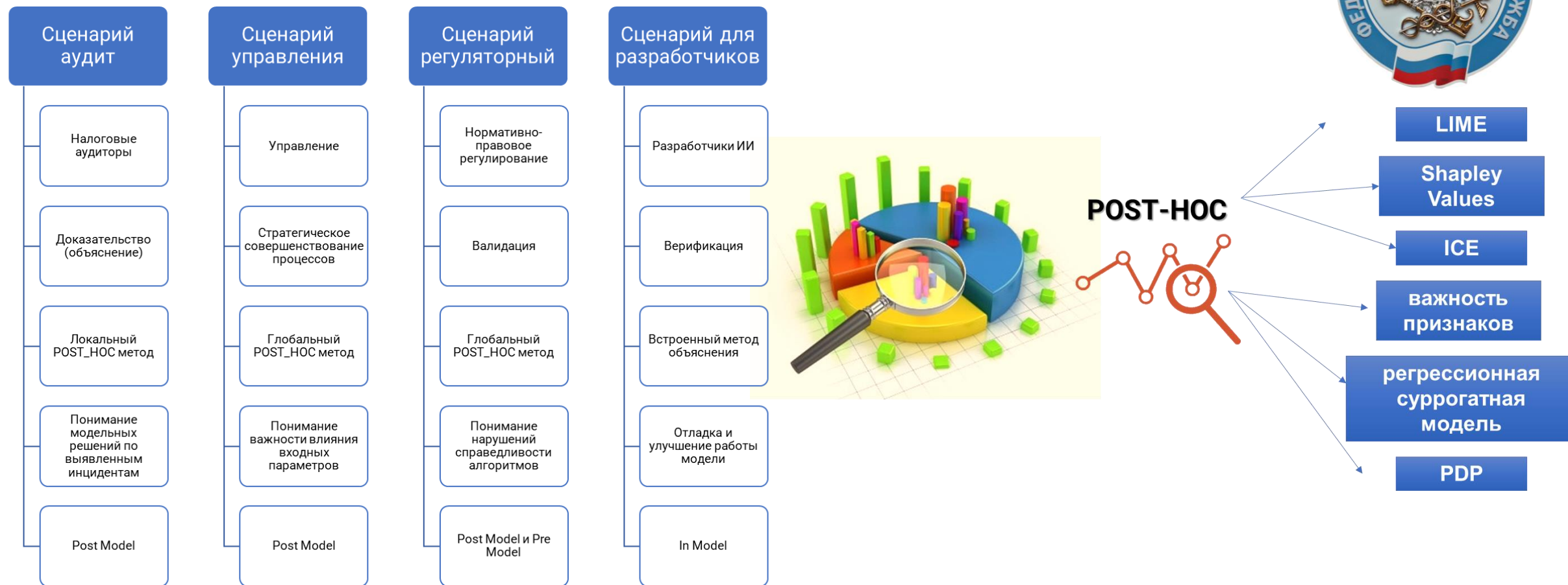
Подходы на основе объяснимого ИИ



Я осознаю почему ИИ принял такое решение
Я понимаю, почему исключены другие варианты
Я знаю, когда ИИ добивается результата, а когда терпит неудачу
Я взвешиваю риски и отдаю отчет, когда ИИ можно доверять
Я знаю, почему и где ИИ ошибается для исправления ошибки

Роль и место интерпретируемого ИИ в государственном управлении

ИСППР модуль с предиктивным анализом целевых организаций для налогового аудита



• Принципы интерпретируемого ИИ

Национальный институт стандартов и технологий (NIST) опубликовал в августе 2020 года первый проект перечня принципов объяснимого искусственного интеллекта (XAI)

ОБЪЯСНЕНИЕ (EXPLANATION)

ТОЧНОСТЬ ОБЪЯСНЕНИЯ
(EXPLANATION ACCURACY)

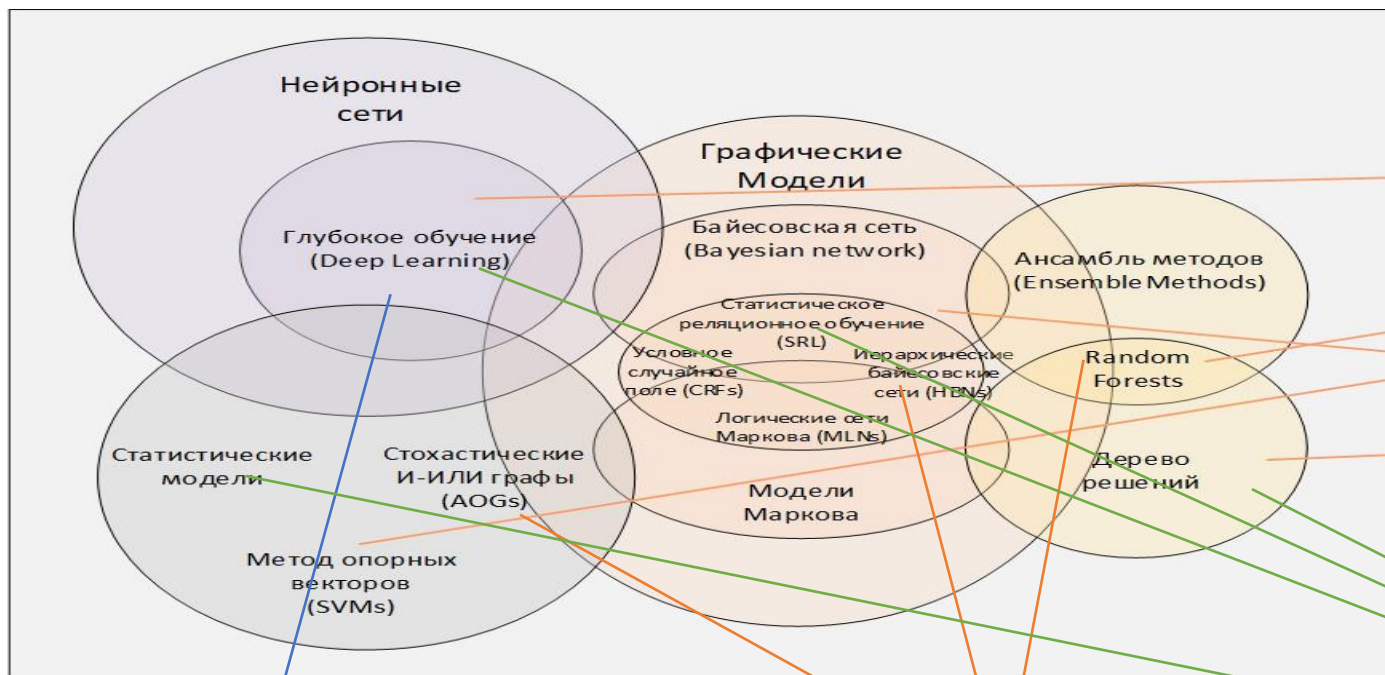
ОСМЫСЛЕННОСТЬ (MEANINGFUL)

ПРЕДЕЛЫ ЗНАНИЙ (KNOWLEDGE LIMITS)

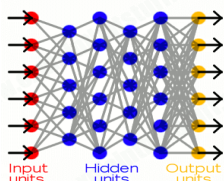
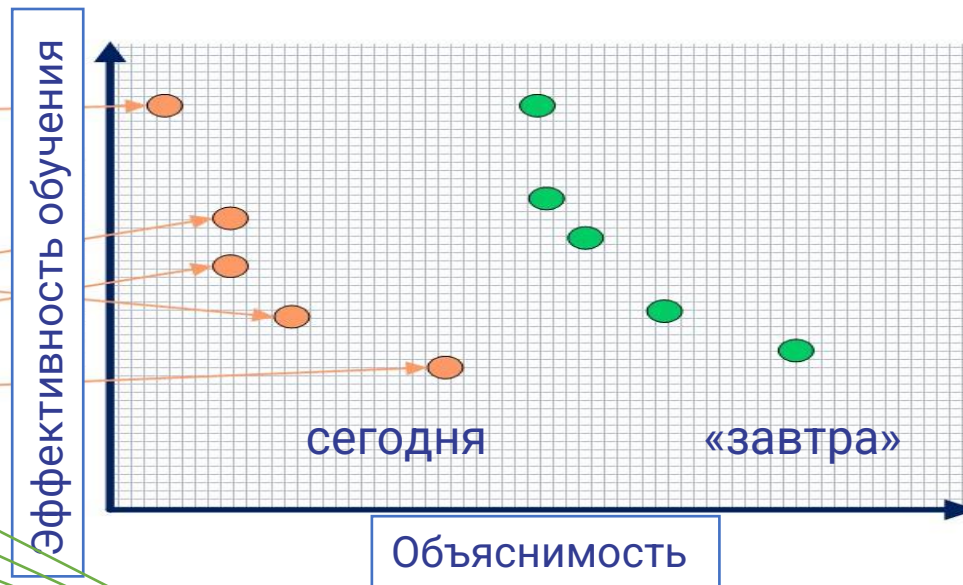


Подходы интерпретируемого ИИ

Текущие методы обучения ИИ

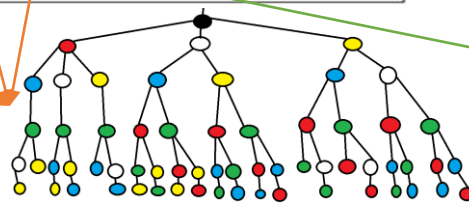


Ожидание по развитию объяснимого интеллекта



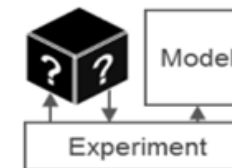
Глубокое объяснение

Модифицированные методы глубокого обучения объяснимым признакам



Интерпретируемые модели

Методы изучения более структурированных, причинно-следственных моделей



Модель индукции

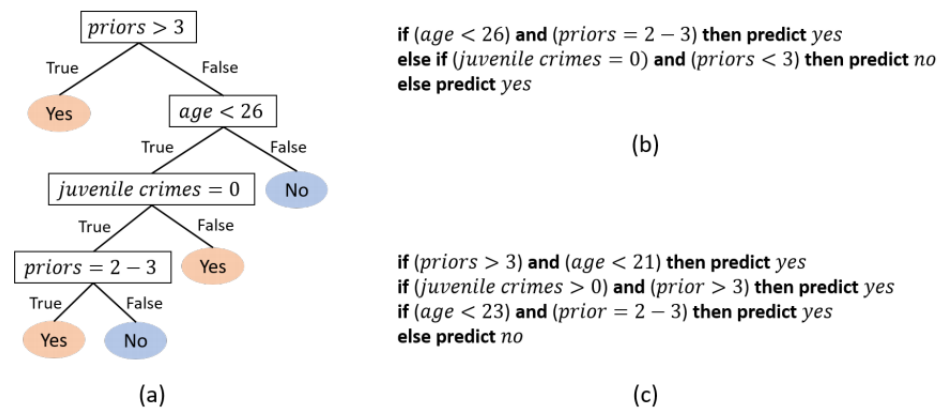
Методы вывода объяснимого Решения из любой модели в виде черного ящика

-
- # Методы интерпретируемой классификации
-

ЛИНЕЙНЫЕ МОДЕЛИ КЛАССИФИКАЦИИ

Скоринговые \ оценочные системы - оценка рисков

Цель – получение линейного решающего правила с небольшим числом термов.



Patient screens positive for obstructive sleep apnea if Score > 1			
1.	age ≥ 60	4 points
2.	hypertension	4 points	+
3.	body mass index ≥ 30	2 points	+
4.	body mass index ≥ 40	2 points	+
5.	female	-6 points	+
Add points from row 1-6		Score	=

ДЕРЕВЬЯ РЕШЕНИЙ \ РЕШАЮЩИЕ СПИСКИ \ РЕШАЮЩИЕ МНОЖЕСТВА ПРАВИЛ

Разреженные логические модели - определения значений показателей надежности

Цель – получение простого решающего правила, понятного человеку
(в терминах логики и правил вида «если-то»)

• Логический анализ данных (интерпретируемое машинное обучение)

Выявление частичных (нечетких) логических закономерностей (оптимальных паттернов данных) в больших массивах численных данных

Обнаружение комбинаций значимых признаков

Выявление наиболее типичных наблюдений класса

Обработка скудных исходных данных
(пропуски в данных, малое число объектов)

Обнаружение новых классов наблюдений

Определение «выбросов» в данных

Построение компактного (разреженного) классификатора

Обоснование и интерпретация решений

Разреженные логические модели

логический
анализ данных
(LAD)

метод
покрывающих
множеств
(SCM)

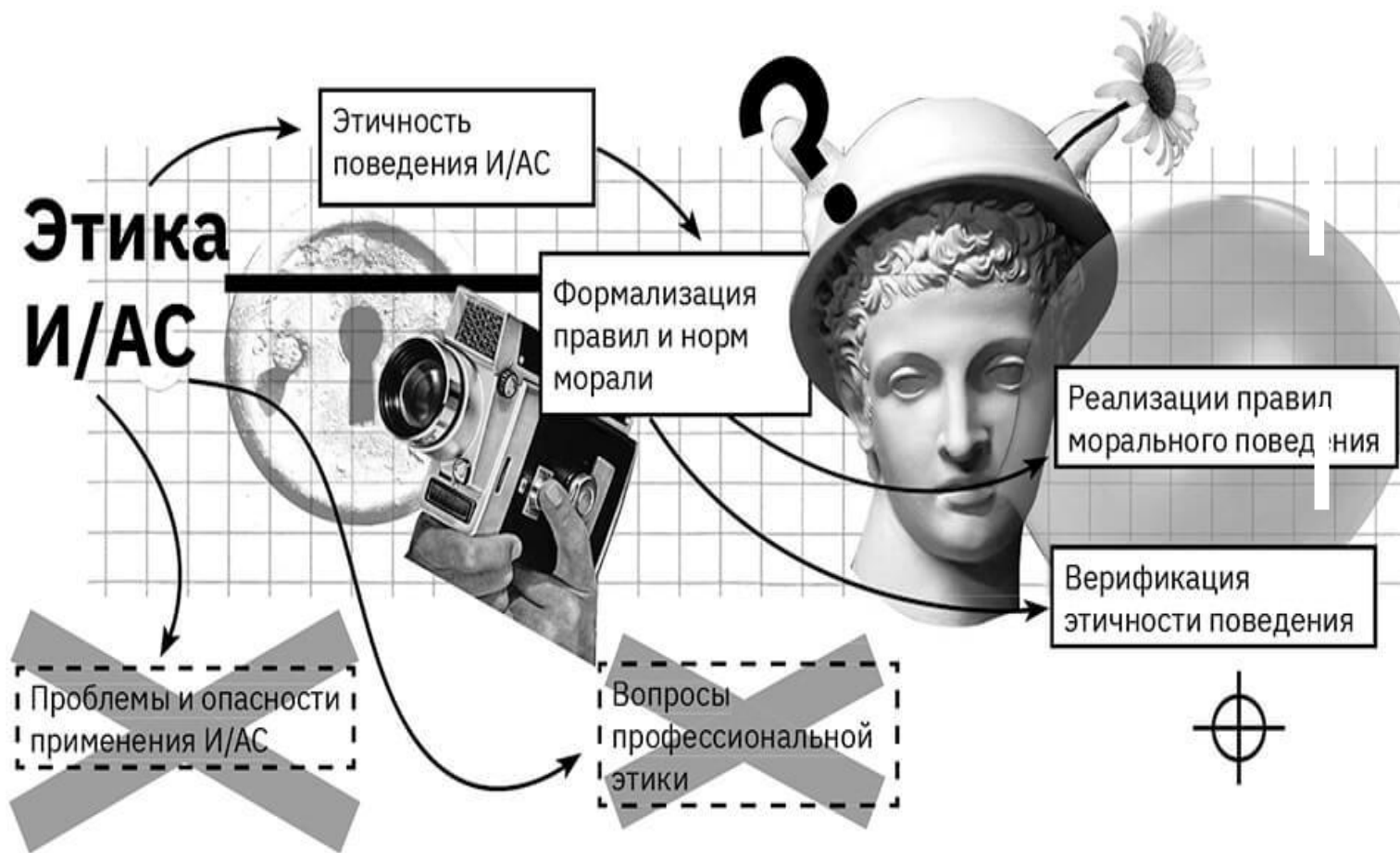
теория тестов
(TT)

теория грубых
множеств
(RS)

анализ
формальных
понятий
(FCA)

Логические модели, состоящие из логических утверждений, включающих предложения «если-то», «или» и «и», являются одними из самых популярных алгоритмов для интерпретируемого машинного обучения, поскольку их утверждения предоставляют понятные человеку причины для каждого прогноза.

Этические вопросы и ограничения ИИ







ОБУЧЕНИЕ МАШИН ВОЗМОЖНО ТОЛЬКО
НА ОСНОВЕ МАССИВА ДАННЫХ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ
ОГРАНИЧЕНЫ КОНКРЕТНЫМ ВИДОМ
ДЕЯТЕЛЬНОСТИ

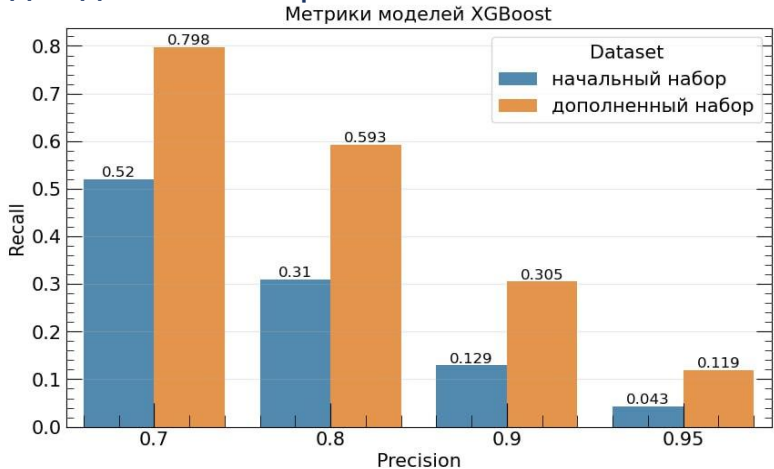
ИНТЕЛЛЕКТУАЛЬНЫЕ МАШИНЫ НЕ
ЯВЛЯЮТСЯ АВТОНОМНЫМИ

Идентификация юр. лиц с недостоверным адресом по данным из ЕГРЮЛ

Разработанное программное решение поддерживает следующий функционал:

-  Загрузка сведений из ЕГРЮЛ в формате CSV
-  Выявление юр. лиц, указавших недостоверный адрес, с определённой пользователем точностью (precision)
-  Сортировка и фильтрация по колонкам с результатами классификации в таблице внутри приложения
-  Сохранение таблицы с результатами для дальнейшей работы с ней

Соответствие значений точности (precision) и охвата (recall) модели на исходных и дополненных данных



Вычисление вклада признаков в вероятность наличия недостоверных данных методом Шепли (SHAP)



Загрузка данных

График PR кривой

Таблица с классификацией

Загрузить данные для предсказания в формате CSV

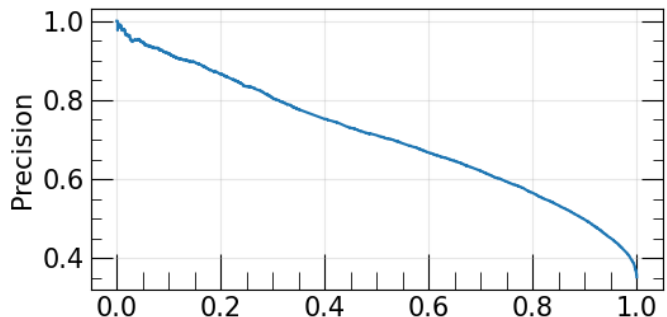
Построить PR кривую

Загрузка данных

График PR кривой

Таблица с классификацией

auc = 0.7091



Введите желаемое значение precision от 0 до 1

0.8

Выполнить классификацию

Фильт...

Фильтр ...

Фильтр prediction...

Фильтр label_pr...

ID	REGION	prediction_proba	label_predicted
6413605	36	0,98027	1
6460918	48	0,965433	1
6440118	77	0,964502	1
6880180	74	0,955973	1
6685513	56	0,939726	1
6570854	16	0,933069	1
6474087	77	0,92834	1
6555925	77	0,927979	1

Предыдущая страница

1...1000 / 1000

Следующая страница

Сохранить таблицу



**ФНС
ЛАБ**



НОЦ ФНС РОССИИ
И МГТУ ИМ. Н.Э. БАУМАНА

Спасибо за внимание!

