



СБЕР
КИБЕР
БЕЗОПАСНОСТЬ

Мировой опыт доверенного искусственного интеллекта



Опыт саморегулирования на примере Big Tech



Этическое саморегулирование (2023 год)

- Принято крупнейшими разработчиками GenAI в диалоге с государством
- Призвано обеспечить кибербезопасность ИИ-моделей
- Присоединились Amazon, Google, Meta*, Microsoft и другие

Рекомендованные практики Google:

- Человекоцентричный дизайн
- Тщательный отбор и проверка данных
- Понимание ограничений модели
- Тестирование и мониторинг

Основы ответственного ИИ Meta*:

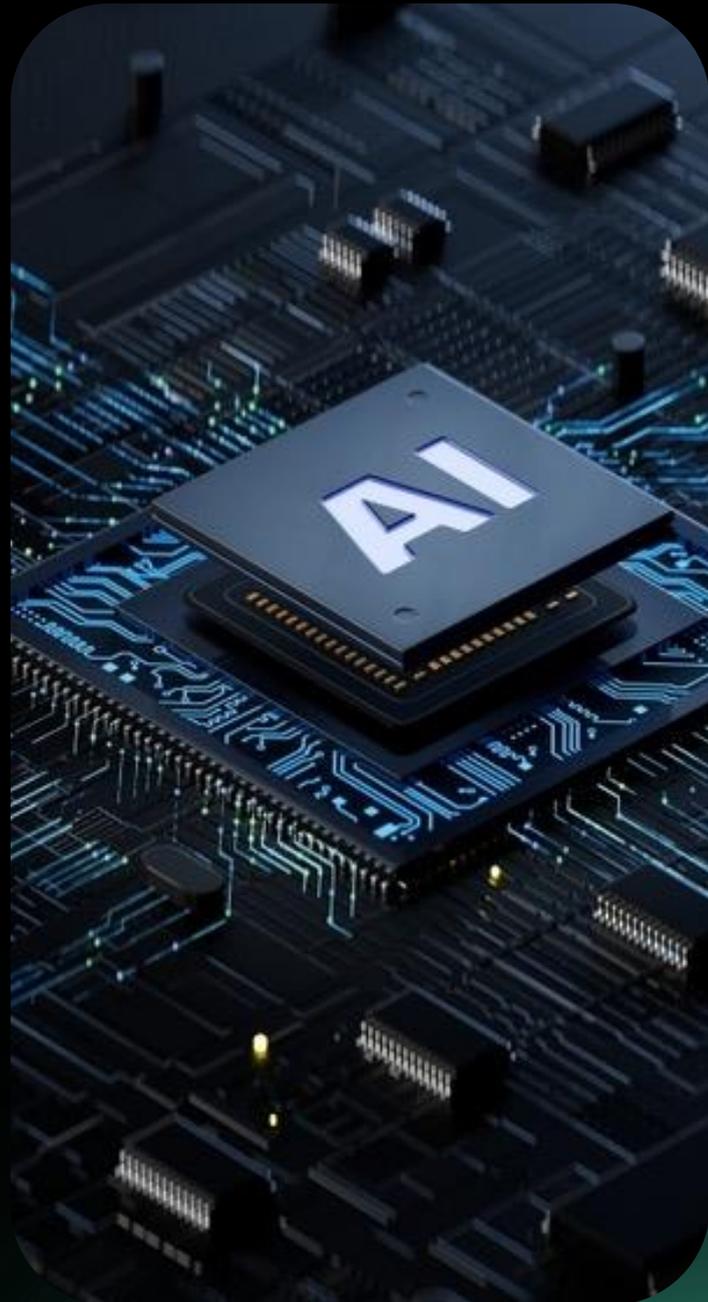
- Конфиденциальность и безопасность
- Справедливость и инклюзивность
- Надежность
- Прозрачность и контроль
- Учет и управление

Построение доверенного ИИ в технологических компаниях

- Microsoft AI Transparency Report
- Google Responsible AI practices
- Meta* Responsible AI

* Признана экстремистской организацией и запрещена в РФ

Подходы к регулированию ИИ



Ограничительный

Приоритет ограничительным мерам в отношении ИИ

Евросоюз



Проинновационный

Развитие и саморегулирование ИИ

Сингапур, Великобритания, Индия и др.



Гибридный

Точечные ограничения и саморегулирование

США, Китай

Ограничительный подход



ЕВРОСОЮЗ

2018 год

Вступил в Силу Общий регламент по защите данных (GDPR)

- Технические меры защиты персональных данных
- Высокие оборотные штрафы за несоблюдение регламента



2023 год

Вступил в Силу Закон о цифровых услугах (DSA)

- Регулирование деятельности технологических гигантов
- Ответственность Big Tech за пользовательский контент



2024 год

Европарламент проголосовал за принятие закона о ИИ (AI ACT)

- Запрет на некоторые системы ИИ
- Обязательства для разработчиков
- Работает в связке с GDPR и DSA

Проинновационный подход свойственен догоняющим странам



Великобритания

- Принципо-ориентированный подход
- Регуляторные песочницы (госорганы + бизнес)



Индия

- Создание благоприятной среды для развития ИИ
- Выделение ключевых секторов применения ИИ
- Международные усилия по доверенному ИИ



Сингапур

- Запущена платформа AI Verify для доверенного ИИ
- Созданы правовые условия для тестирования беспилотных автомобилей



Израиль

- Мягкие инструменты регулирования ИИ – этика, стратегические и разрешительные акты
- Ведутся консультации с бизнесом по вопросам регулирования ИИ

Гибридный подход



США

Саморегулирование и точечные ограничения

- Добровольные этические обязательства
- Отчеты корпораций о построении доверенного ИИ (см. слайд 2)
- Запретительное регулирование на уровне штатов

Указ о безопасном и надежном искусственном интеллекте (2013 год)

- Продвижение отраслевых стандартов
- Совместная работа с бизнесом и наукой

Фреймворк AI RMF (NIST)

Характеристики доверенного ИИ:

- Валидность и надежность
- Безопасность
- Объяснимость и интерпретируемость
- Непредвзятость
- Конфиденциальность
- Подотчетность и прозрачность

Гибридный подход



КИТАЙ

Ограничительное регулирование отдельных сфер

- Запрет использования дипфейков в ряде направлений (2023)
- Меры по управлению продуктами на основе генеративного ИИ (2023)
- Регулирование рекомендательных алгоритмов (2021)

Этический кодекс для ИИ

- Содействие благополучию людей
- Повышение справедливости
- Защита конфиденциальности и безопасности
- Обеспечение управляемости и надежности
- Повышение эффективности
- Внимание этическим вопросам

Технические меры для доверенного ИИ

- Разработка моделей, объясняющих решения ИИ
- Проверка производительности и выявление ошибок
- Шифрование для безопасного хранения и передачи данных
- Алгоритмы, не допускающие дискриминации

Проект закона об ИИ

- Приоритет развитию отрасли ИИ перед защитой пользователей
- Ограничительные меры в отдельных областях ИИ
- Допускает участие третьих сторон в тестировании ИИ-моделей

Построение доверенного ИИ



РОССИЯ

В России внедряются лучшие мировые практики развития ИИ

- Приняты стратегические акты и рекомендации
(Указ Президента РФ от 10 октября 2019 г. N 490 "О развитии искусственного интеллекта в Российской Федерации",
Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.)
- Принято законодательство об экспериментальных правовых режимах
(123-ФЗ, 258-ФЗ)
- Сняты отдельные отраслевые барьеры
(тестирование беспилотного транспорта, использование ИИ в медицине)
- Принят Национальный кодекс этики ИИ (2021 г.)

Готов ответить на ваши вопросы



СБЕР
КИБЕР
БЕЗОПАСНОСТЬ



Петр Хенкин

Исполнительный директор,
Департамент кибербезопасности

PVKhenkin@sberbank.ru