KPMG

DATA vs. AI Governance

Konstantin Aushev

—

September 6, 2024

Первые штрафы за недостатки в управлении данными

Проект Oncology AI, получив 5 млрд долл. инвестиций, был закрыт после обнаружения систематического некорректного диагностирования пациентов

Крупная компания в сфере недвижимости анонсировала списание 304 млн долл. из-за завышенной оценки при покупке объектов, обусловленной некорректными данными, использованными их интеллектуальными алгоритмами

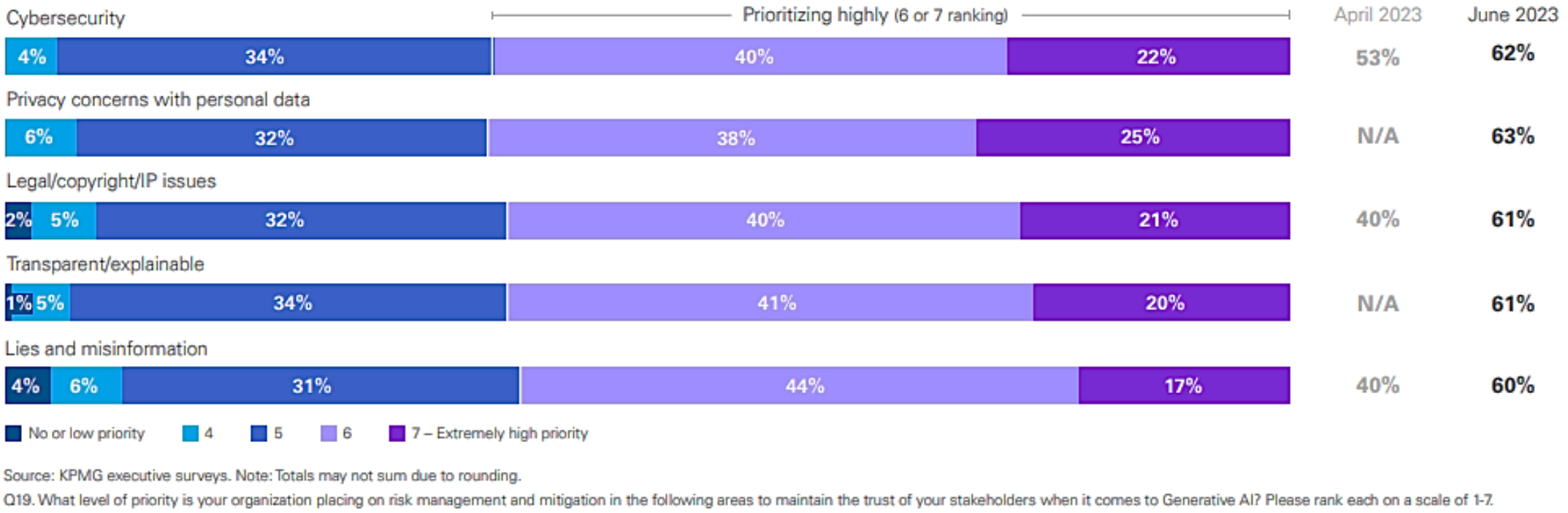
Компания в сфере e-коммерции была обвинена в использовании сексистского рекрутингового AI-инструмента

Недостаток в скоринговой модели привел к 80% необоснованных отказов по расовой принадлежности

Медиакомпания была оштрафована на 5 млрд долл. за недостатки в процессах управления данными и защиты приватности

Кибератаки привели к сбоям в работе ИИ-моделей по диагностике заболеваний

60+% руководителей связывают с GenAI увеличение рисков

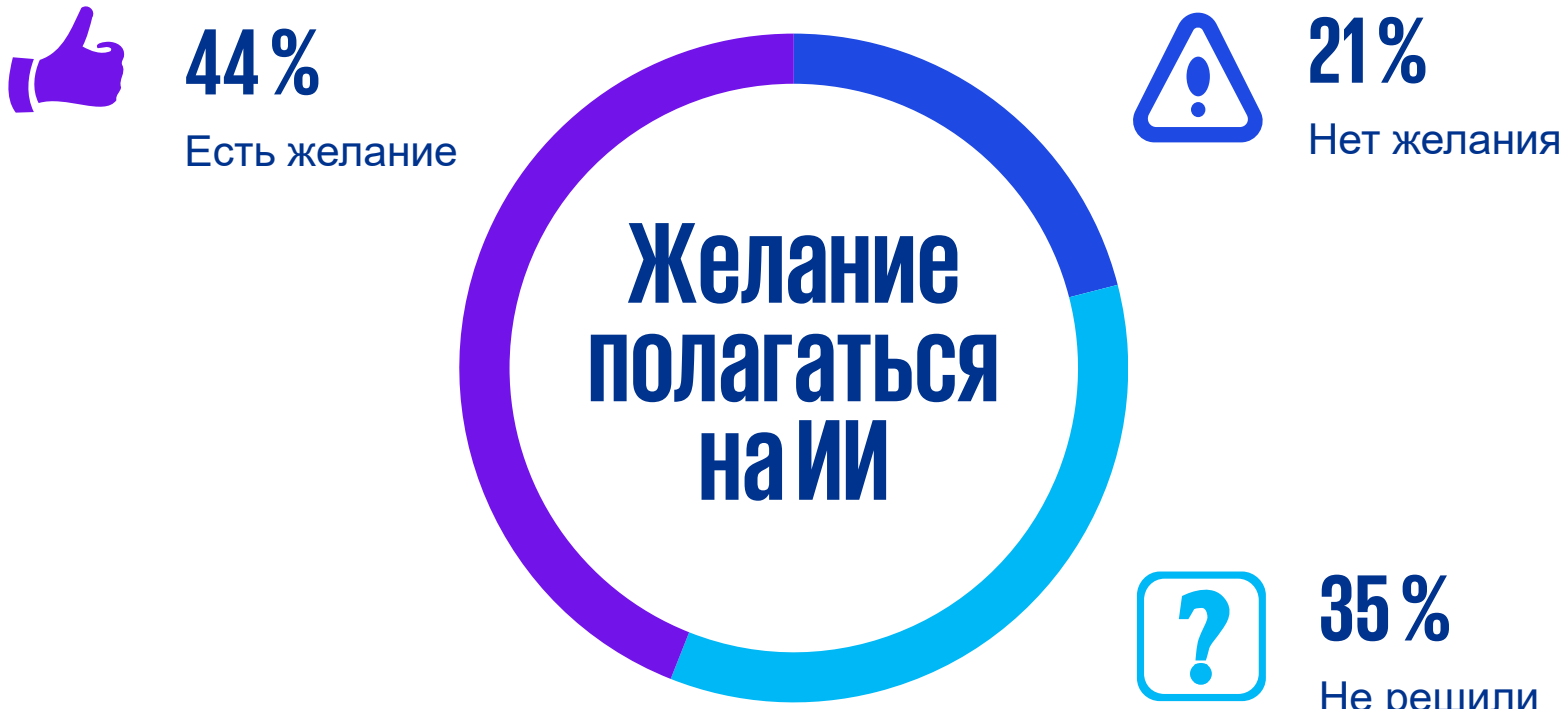


Source: KPMG executive surveys. Note: Totals may not sum due to rounding.

Q19. What level of priority is your organization placing on risk management and mitigation in the following areas to maintain the trust of your stakeholders when it comes to Generative AI? Please rank each on a scale of 1-7.

Источник: KPMG Trusted AI Global Insights.

44 % готовы полагаться на выводы ИИ



Источник: KPMG Trusted AI Global Insights.

Но 63 % не уверены, что готовы делиться данными с ИИ

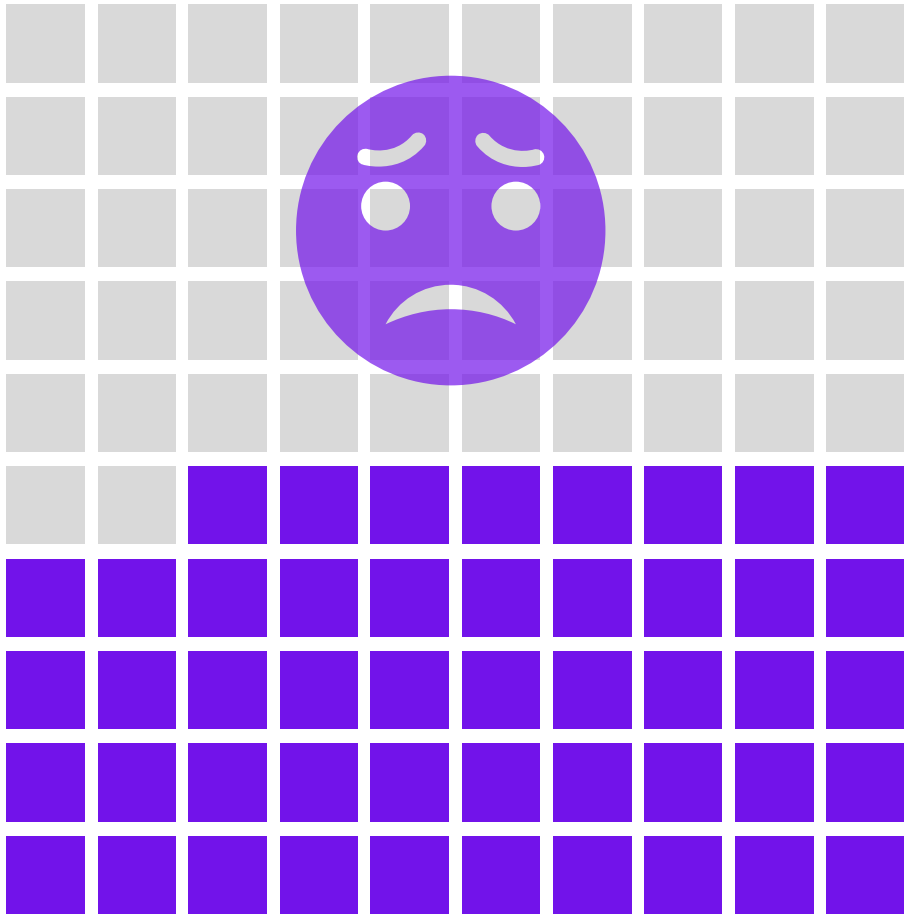
37 %
Есть желание

33 %
Нет желания

30 %
Не решили

Источник: KPMG Trusted AI Global Insights.

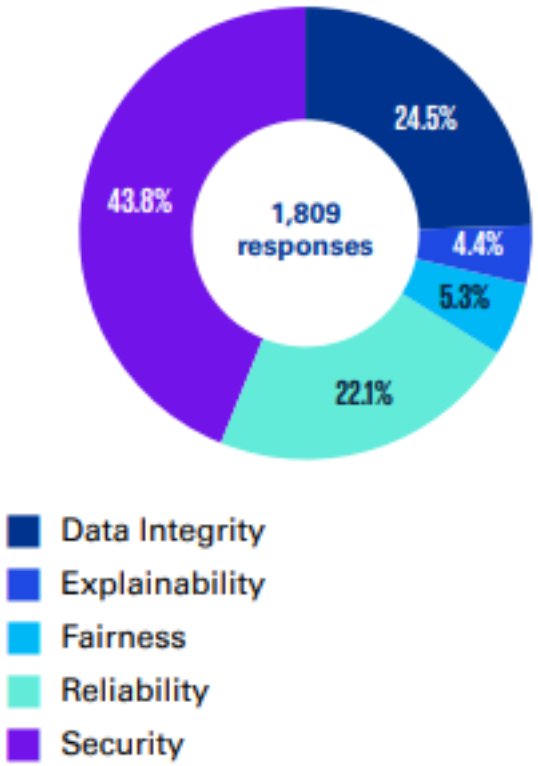
48 % испытывают страх или беспокойство от ИИ



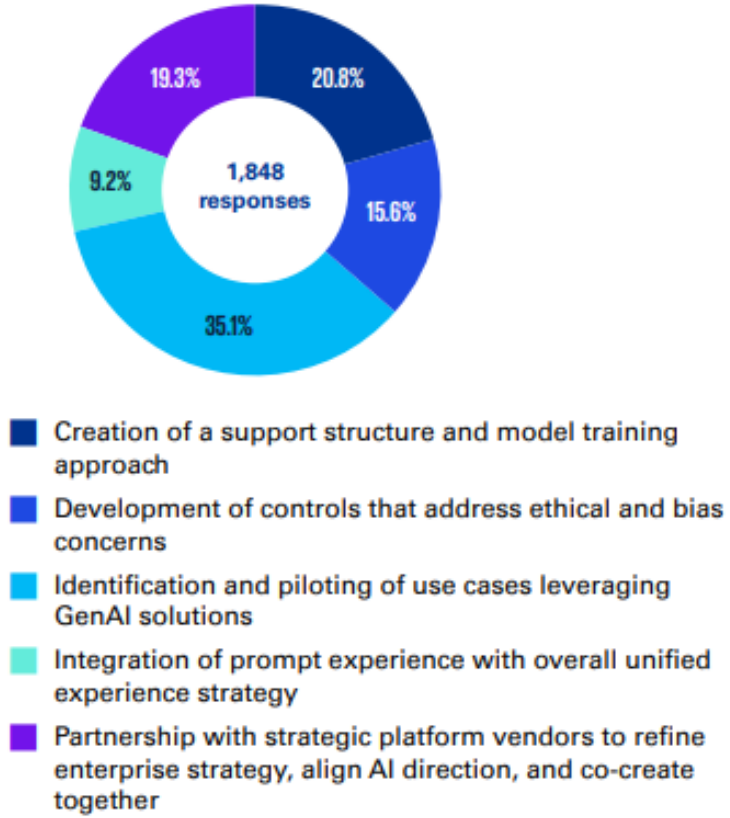
Источник: KPMG Trusted AI Global Insights.

Вопросов больше, чем ответов

Какие риски, связанные с данными и моделями ИИ, наиболее применимы к вашей организации?

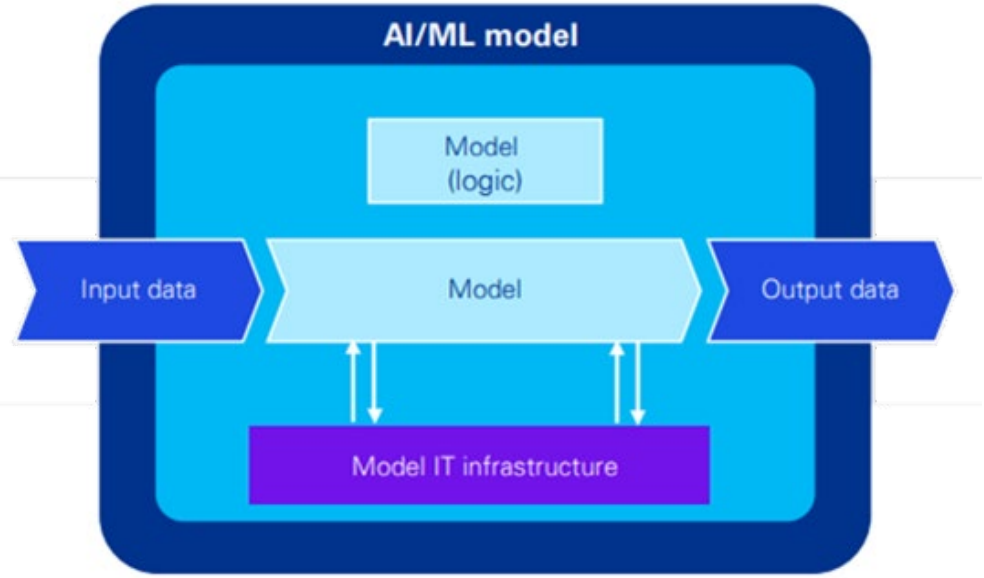


На каком этапе внедрения GenAI вашей организации понадобится больше помощи?



Треть говорят, что могут снизить риски применения ИИ, но по факту не знают, с чего начинать

1. **Риски неэффективного управления и определения контекста моделей**
2. **Риски некачественных входных данных**
3. **Риски искажения выходных данных**
4. **Инфраструктурные риски и риски логики модели**



Что отличает GenAI?

1. **Расширенные лингвистические модели, натренированные на сотнях триллионов параметров**

2. **Большое количество преднастроенных на больших данных моделей, способных выдавать результат очень быстро**

3. **Техники «наращивания» данных позволяют увеличивать выборки, улучшать качество данных, улучшать показатели моделей**

4. **Встраивание по умолчанию лингвистических движков, что упрощает взаимодействие с моделью**

Сгенерировано с помощью MS Designer.

4 зоны рисков адаптации GenAI, к которым нужно подходить с точки зрения управления данными

Устойчивость

Риск чрезмерного потребления энергии на хранение и обработку данных

> Контролировать количество собираемых данных

Транспарентность

Риск потери доверия стейкхолдеров

> Информировать субъекты о целях сбора и способах обработки данных

Защита персональных данных

Риск нарушения прав субъектов, чьи данные собираются для анализа

> Маскировать собираемые данные, минимизировать сбор данных

Честность

Риск сбора нерепрезентативных данных

> Определить и следовать принципам сбора и обработки данных

Trusted AI – цель номер 1

Хорошо понимая вопросы доверия и этики в области ИИ, регуляторные и технологические вызовы, в KPMG мы разработали фреймворк для управления ИИ в бизнесе, нацеленный на практическое применение здесь и сейчас.

Честность
Исключение смещенных оценок из-за субъективности отдельных людей или их групп

Прозрачность
Обеспечение возможности указания необходимых раскрытий для понимания стейкхолдерами принципов работы алгоритмов ИИ

Объяснимость
Возможность понять, как и почему решения ИИ сгенерировали те или иные рекомендации

Ответственность
Встроенные механизмы мониторинга на всем цикле работы ИИ для управления рисками и обеспечения регуляторного соответствия

Защита данных
Механизмы защиты от неавторизованного доступа, злоумышленников, подложных данных, повреждения данных и систем

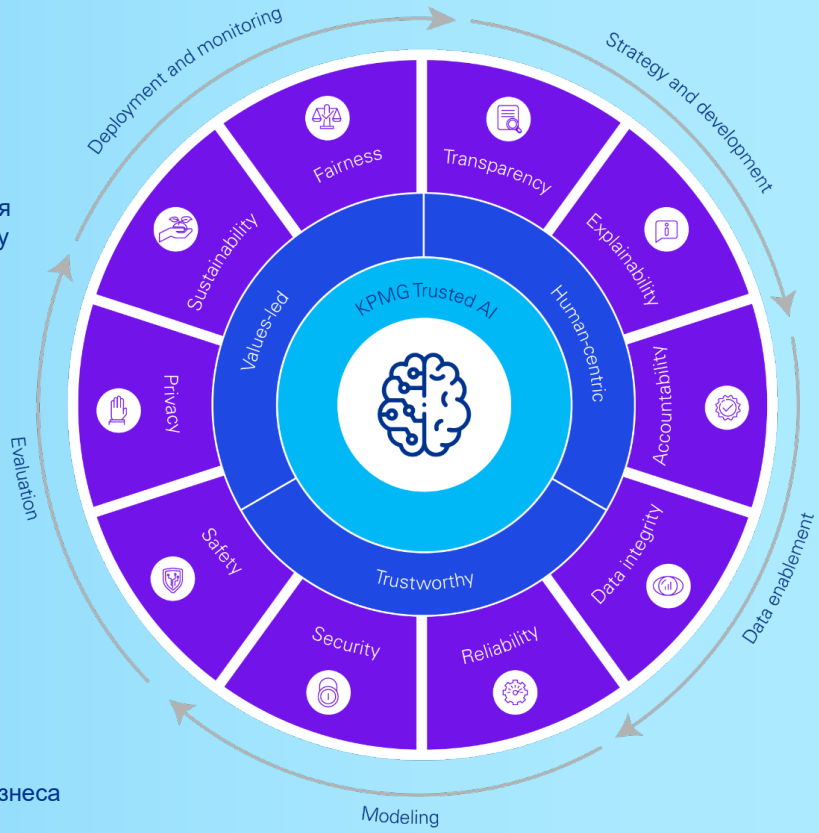
Приватность
Обеспечение соответствия локальному и международному законодательству в области персональных данных

Устойчивость
Оптимизация решений ИИ с целью уменьшения негативного воздействия на окружающую среду

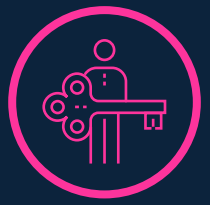
Целостность данных
Обеспечение качества данных, управляемости данных, доверия к данным

Надёжность
Обеспечение заданного уровня точности и консистентности в работе ИИ

Личная безопасность
Механизмы защиты в решениях ИИ от потенциального вреда для людей и активов бизнеса

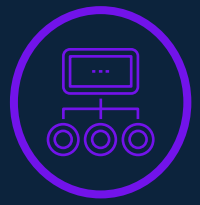


Необходимость инвестиций в AI Governance сегодня



Ключевые компоненты AI Governance

- Процессы выравнивания взаимодействий внутри организации и обеспечения следования установленным принципам Trusted AI
- Операционная модель и связь с поддерживающими процессами и технологиями
- Четкое определение ролей и ответственности для всех стейкхолдеров, вовлеченных в инициативы ИИ
- Адаптация KPI для гарантии непрерывного развития ИИ в организации



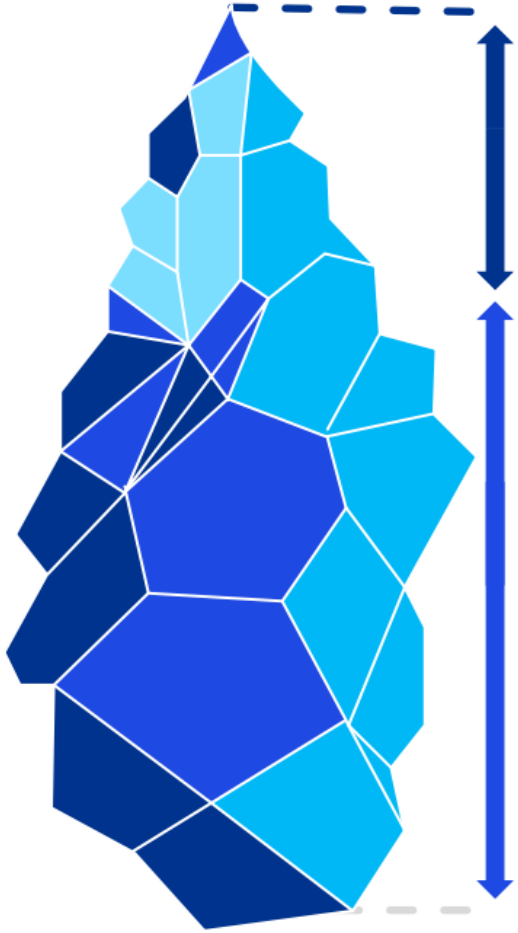
Мощная модель AI governance…

- Гарантирует, что развитие ИИ в организации будет идти в соответствии с ее целями
- Гармонизирует различные инициативы внутри организации с общей стратегии, приоритетами и имеющимися ресурсами
- Ускорит реализацию инноваций, появление новых идей и их прототипирование
- Позволит четко отслеживать прогресс и гибко регулировать необходимые инвестиции в развитие ИИ в организации

Типовые триггеры наших клиентов

- ✓ Нам необходимо **защититься** от финансовых и репутационных рисков
- ✓ Нам нужно **увеличить доверие клиентов** (внешних, внутренних)
- ✓ Нам нужно **установить четкие границы ответственности**
- ✓ Нам нужно **доказать ценность** дальнейшего развития ИИ в масштабах всей организации
- ✓ Нам нужно **обеспечить соответствие** локальным и международным законам в области ИИ
- ✓ Нам нужно **защитить наши модели** от атак

Из практики – о чем задуматься уже сейчас



Очевидные задачи

- Определить видение и целевую архитектуру
- Отслеживать затраты и эффекты на эксперименты
- Вовлекать бизнес-пользователей
- Рассматривать ИИ-инициативы как полноценную программу проектов с соответствующими механизмами управления

Менее очевидные критические задачи

- Внедрить механизмы отслеживания и анализа ошибок моделей
- Оценить риски, связанные с поставщиками каждой отдельной услуги и технологии
- Вовлекать комплаенс-службы на каждом шаге проектирования
- Прорабатывать интеграции с другими решениями в организации на самом раннем этапе программ внедрения ИИ-решений
- Устраивать кросс-функциональные сессии команд с выстроенными подходами к управлению данными
- Внедрять изменения посредством постоянных коммуникаций
- Непрерывно работать над улучшением качества данных

Константин Аушев

Партнер,
руководитель
Технологической
практики

kaushev@kpmg.kz

kpmg.kz