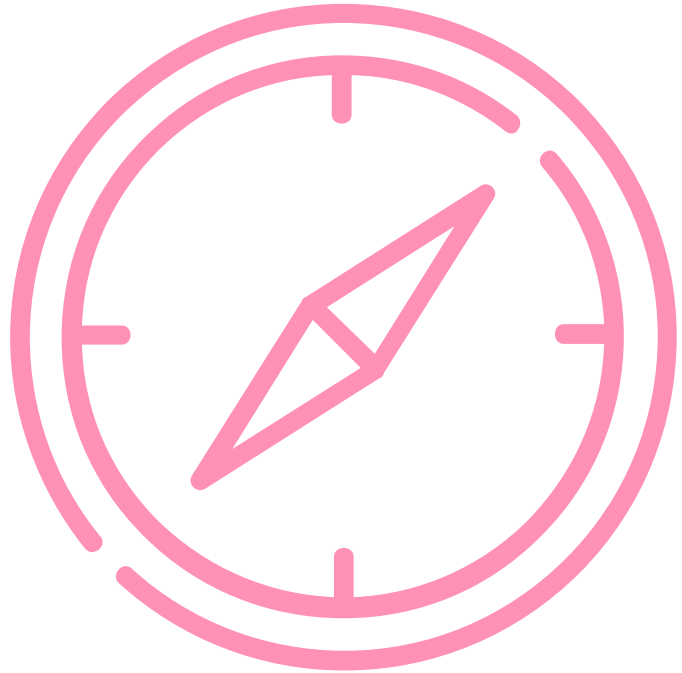
Особенности применения ИИ в финансовом секторе

Товстолип Александр

Ассоциация ФинТех

Содержание

- Что такое ИИ
- ИИ в финсекторе
- Риски и угрозы ИИ
- Безопасность ИИ и MLSecOps
- Что важно для финансовой отрасли?



Ассоциация ФинТех

АССОЦИАЦИЯ ФИНТЕХ

13 ЧЛЕНЫ АФТ

Банк России
ВТБ
СБЕР
Альфа Банк
ГАЗПРОМБАНК
НСПК
ПСБ
ВСК страховой дом
РСХБ | ЦИФРА
MOEX GROUP
Т БАНК
Райффайзен БАНК
ДОМ РФ

33 АССОЦИИРОВАННЫЕ ЧЛЕНЫ

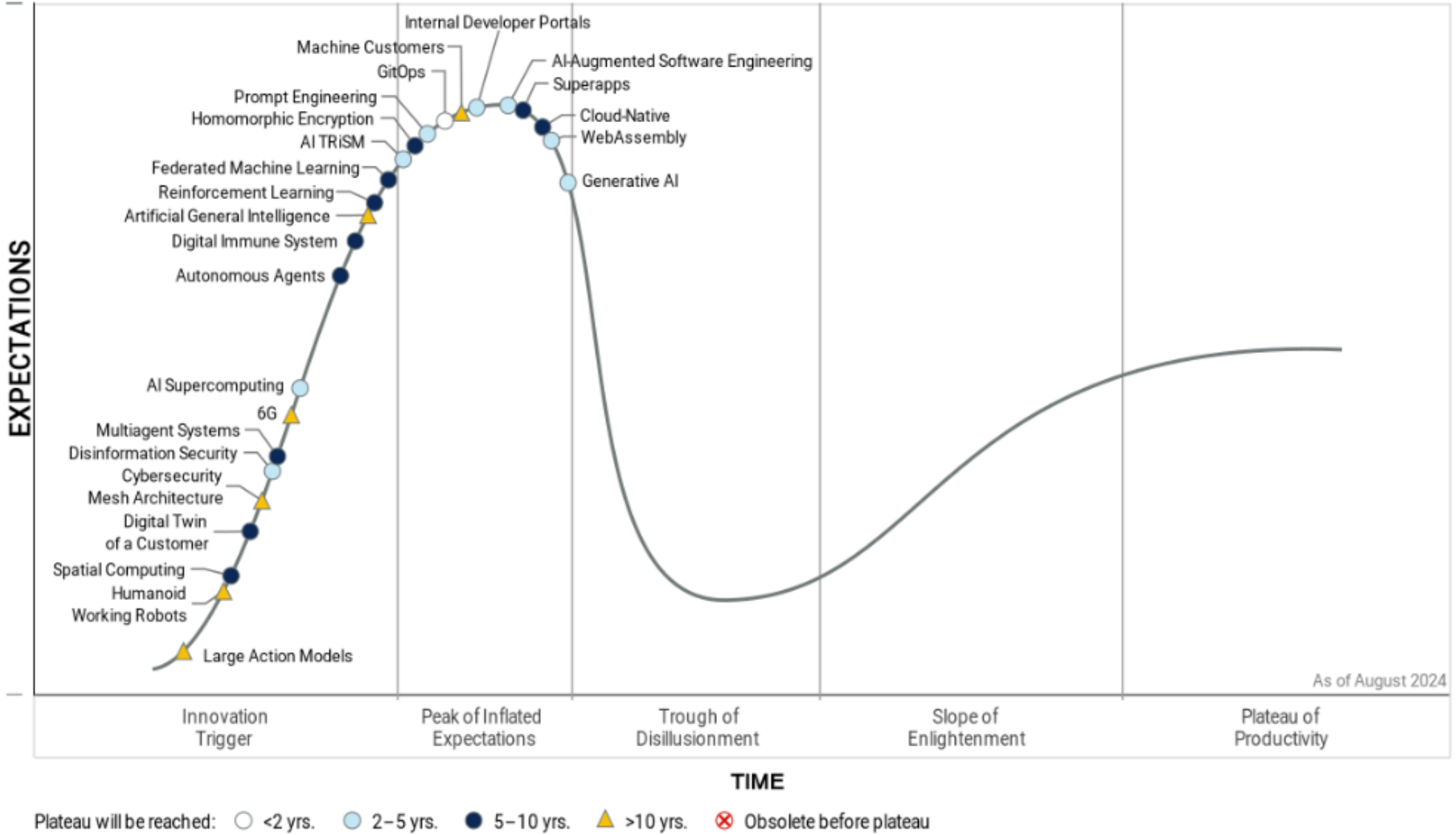
СОВКОМБАНК
base alt
РНКО ПЛАТЕЖНЫЙ ЦЕНТР
ART искусство прибыли
МКБ
Яндекс
МТС
НСИС Национальная Страховая Информационная Система
ПОЧТА БАНК
Ростелеком
Avito
РЕСО гарантия
Ак Барс Банк
DIASOFT всё по-настоящему
БАНК ЦЕНТР-ИНВЕСТ
РЕДСОФТ
КАПИТАЛ LIFE страхование жизни
отпбанк
АЛЬФА СТРАХОВАНИЕ
K2TEX
ДОХОДЪ управляющая компания
сравни
Кредит Европа Банк
синара банк
TI
билайн бизнес
БПС инновационные решения
ВБРР
АСТРА
точка банк для предпринимателей и предприятий
SWORDFISH SECURITY
YADRO
FLEX software systems

Что такое искусственный интеллект?

Комплекс технологических решений, позволяющих имитировать когнитивные функции человека (включая самообучение, поиск решений без заранее заданного алгоритма и создание контента) и получать при выполнении задач результаты, сопоставимые с результатами интеллектуальной деятельности человека.

Набор данных, алгоритмов, программного обеспечения, систем и экспертизы.

Gartner 2024 Hype Cycle for Emerging Technologies Highlights Developer Productivity, Total Experience, AI and Security



Рынок ИИ растет

650
млрд ₽

Общий объем рынка ИИ в России по состоянию на 2023 год

11,2
трлн ₽

Может дать внедрение ИИ в ключевых отраслях экономики до 2030 года

>25% - ежегодный темп роста рынка

6% роста ВВП России

32,5
млрд ₽

Выделено финансирования на федеральный проект «Искусственный интеллект» до 2024 года

15-40%
мирового ВВП

Будет обеспечено за счет внедрения ИИ до 2030 года по прогнозу McKinsey

ИИ В ФИНСЕКТОРЕ

ИИ в функциональных процессах организации

Фронт. функции	**Продажи**	**Клиентское обслуживание**	**Маркетинг**
Обслуживающие функции (Core)	**Управление клиентскими данными**	**Управление взаимоотношениями с клиентами**	**Управление рисками**
	Управление продуктами	**Обслуживание бизнес операции, операционная деятельность**	Финансовый мониторинг и Compliance
	Карточный бизнес	Управление каналами дистрибуции	Управление партнерами	Актуарные расчеты	Управление убытками
	Управление просроченной задолженностью	Взаимодействие с регулятором	Взаимодействие с гос. органами
	Учет: налоговый	продуктовый	бухгалтерский
Обеспечивающие функции	Управление внутрихозяйственной деятельностью	Управление финансами	**Делопроизводство и ЭДО**
	Управление ИТ	**Управление человеческим капиталом (HR)**	**Отчетность**
	Хранение и использование данных	**Юридическое сопровождение**	Казначейство
	Информационная безопасность		

Применение ИИ в процессах, которые охватывают выделенные функциональные процессы функциональной архитектуры финансовой организации

Что волнует банки?

1

НОВЫЕ ВОЗНИКАЮЩИЕ РИСКИ

- изменяются ландшафты организаций.
- появляются новые технологии, недостаточно изученные и недостаточно прозрачные.
- рынок решений по защите не развит.

2

ЭТИКА - СЛОЖНОСТЬ КОНТРОЛЯ ЗА СОБЛЮДЕНИЕМ ЭТИКИ

нет объективных и прозрачных механизмов оценки соответствия ИИ-решений нормам этики.

3

ИСПОЛЬЗОВАНИЕ ИИ В МОШЕННИЧЕСКИХ СХЕМАХ

использование злоумышленниками для обмана граждан (дипфейки, социнженерия)

4

НИЗКОЕ ДОВЕРИЕ К ИИ НАСЕЛЕНИЯ

5

РЕГУЛЯТОРНАЯ НЕОПРЕДЕЛЕННОСТЬ

законодательство и регуляторные рамки в стадии формирования.

РИСКИ И УГРОЗЫ ИИ

Применение ИИ - риски «сегодняшнего дня»

СОЦИАЛЬНАЯ ИНЖЕНЕРИЯ И МОШЕННИЧЕСТВО

- Убедительная дезинформация
- Возможность манипулирования общественным мнением
- Автоматизация и масштабирование схем мошенничества и киберпреступлений
- Ускорение подготовки преступников, включая и киберпреступников
- Мошенническая реклама и производство фейков
- Создание фишинговых сайтов

УТЕЧКА КОНФИДЕНЦИАЛЬНОЙ ИНФОРМАЦИИ

- Загрузка в чаты данных, части кода, КИ, ПДн
- Высокая скорость обработки утекшей/похищенной информации
- Использование ИИ для обработки и использования коммерческой информации в преступных целях

АТАКИ НА ИНФРАСТРУКТУРУ И СЕРВИСЫ

- Построение сложных моделей атак
- Повышение качества атак: таргетированный фишинг, фокусные рассылки, сокрытие следов и доказательств, создание вредоносного кода
- Повышение количества атак: генерация спама, возможность использования неограниченного числа точек атак

ПОДМЕНА ИДЕНТИФИКАЦИИ

- Кража личности
- Эмуляция действий человека с целью обхода мер безопасности и систем антифрод

Применение ИИ - риски «завтрашнего дня»

«СЕРАЯ» ЗОНА

- Отсутствие регулирования
- Бесконтрольное внедрение ИИ технологий
- Отсутствие этических ограничений
- Возможность получить полный контроль над действиями пользователя
- Неконтролируемое распространение и передача данных
- Устаревшие или умышленно искаженные рекомендации и выводы
- Нарушение прав интеллектуальной собственности

АТАКИ НА ИНФРАСТРУКТУРУ И СЕРВИСЫ

- Разработка новых типов и моделей атак для получения доступа к критичной инфраструктуре
- Сокращение time-to-market взлома (в тч паролей), возможность генерации ключей
- Использование опции deepfake для целей биометрической идентификации
- Снижение порога входа в технологию позволит дилетантам атаковать системы

РАЗВИТИЕ

- Невозможность закупки передовых графических процессоров (gpu), необходимых для обучения нейросетей
- Ограничена возможность строить сервисы с помощью нейросетей американской компании в РФ: OpenAI закрыла доступ к своим сервисам пользователям и компаниям из России
- ESG: колоссальное количество энергии требуется для развития ИИ

БЕЗОПАСНОСТЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Доверенный ИИ

Данные

исходные и обработанные данные, используемые для обучения модели, могут содержать конфиденциальную информацию

Модели

алгоритмы и параметры модели являются интеллектуальной собственностью и могут быть мишенью для кражи или подделки

Инструменты и инфраструктура

инфраструктура и процессы обучения и развертывания модели могут быть уязвимы к атакам или сбоям

Что делают банки?

1

Все оценивают новые угрозы новых технологий - формируют с той или иной точки зрения модели угроз.

проанализированные нами модели - примерно на 80% похожи.

2

Выстраивают в своих организациях процессы проверки ИИ моделей.

Например, процессы тестирования моделей на защищенность (пентесты, анализ уязвимостей) или выстраивание экосистемы ИИ-агентов, каждый из которых обучен на своем специфическом наборе данных и под какую-то специфическую задачу, в частности на поиск уязвимостей в ИИ-решениях

3

Выстраивают в своих организациях процессы защиты ИИ-решений.

- *выстраивание фильтрующих слоев для анализа запросов (обнаружение промпт-инъекций, запросов на запрещенные темы, обнаружение конфиденциальной информации) и анализа ответов (запрещенная и конфиденциальная информация в ответах)*
- *выстраивание полноценного процесса MLSecOps (LLMSecOps) и обеспечение защиты на каждом этапе этого процесса*

Жизненный цикл безопасного применения ИИ – MLSecOps (включая специфику ГенИИ)

MLSecOps — подход, который объединяет операционные аспекты машинного обучения с вопросами безопасности. Направлен **на снижение рисков**, которые могут принести модели AI/ML/GenAI в организацию.

Что важно для финансовой отрасли?

1 Методика оценки рисков информационной безопасности для решений на основе ИИ

2 Комплексный подход к управлению рисками через призму «технологии – люди – процессы»

3 Решение вопросов оценки соответствия технологий защиты ИИ

4 Проработка вопросов развития мер для защиты ИИ, которые выходят за рамки стандартных механизмов безопасности и используют математические методы

Примером таких мер являются:

- использование синтетических данных для обучения, сгенерированных алгоритмами ИИ;
- обезличивание данных под контролем ИИ;
- использование алгоритмов Differential Privacy для количественной вероятностной характеристики вероятности утечки данных;
- использование ИИ как контролера в процессах с ИИ, который оценивает результаты ИИ на принципы «3Н»: ИИ должен быть полезным (helpful), честным (honest), не вредить (harmless);
- контроль качества данных при помощи фактологических баз знаний;
- ML-методы «забывания» персональных данных после отзыва согласия субъектом ПДн;
- выявление в обучающих наборах конфиденциальной информации и интеллектуальной собственности;
- использование математических методов цифровой подписи контента ИИ, сигнатур использования ИИ для генерации фейковых данных.

Заключение

На что важно обращать внимание:

1. Архитектура безопасности и внедрение процесса MLSecOps
2. Доверие к элементам – данные, модели, инструменты и инфраструктура
3. Повышение экспертизы – специалистов и сообщества
4. Соответствие ИИ этическим нормам
5. Соответствие регуляторным требованиям

СПАСИБО ЗА ВНИМАНИЕ!

Александр Товстолип

Руководитель управления информационной безопасности

tg: @mydzen

a.tovstolip@fintechru.org

Telegram - канал