Искусственный интеллект и машинное обучение для защиты от современных угроз

Константин Саматов

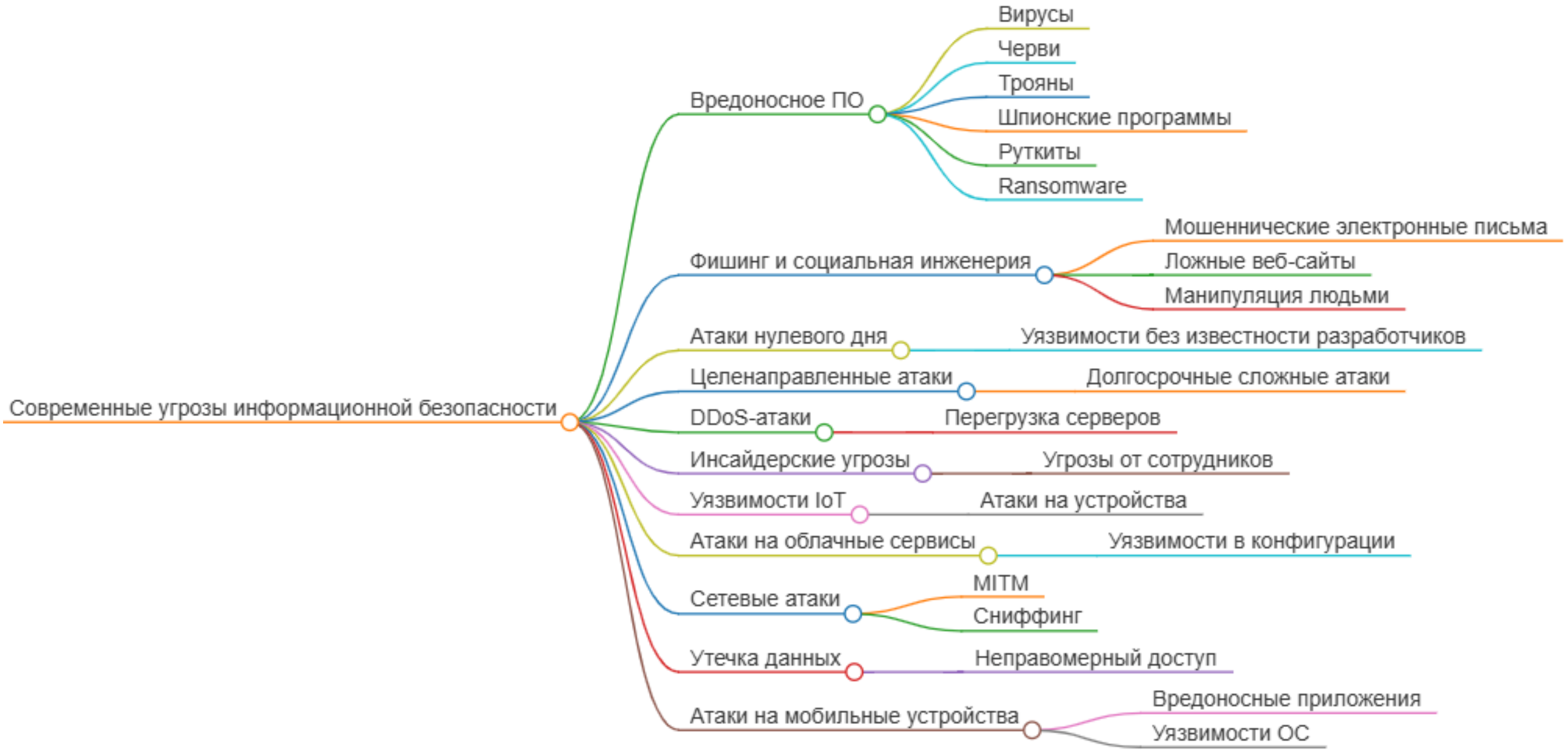
Основные термины связанные с ИИ

Различие между этими терминами заключается в том, что **ИИ** - это более широкое понятие, которое **включает в себя нейросети и машинное обучение**.

Нейросеть - это математическая модель, которая используется для решения задач

Машинное обучение - это класс методов, который используется для обучения компьютерной системы на основе предоставленных данных

Виды современных угрозы

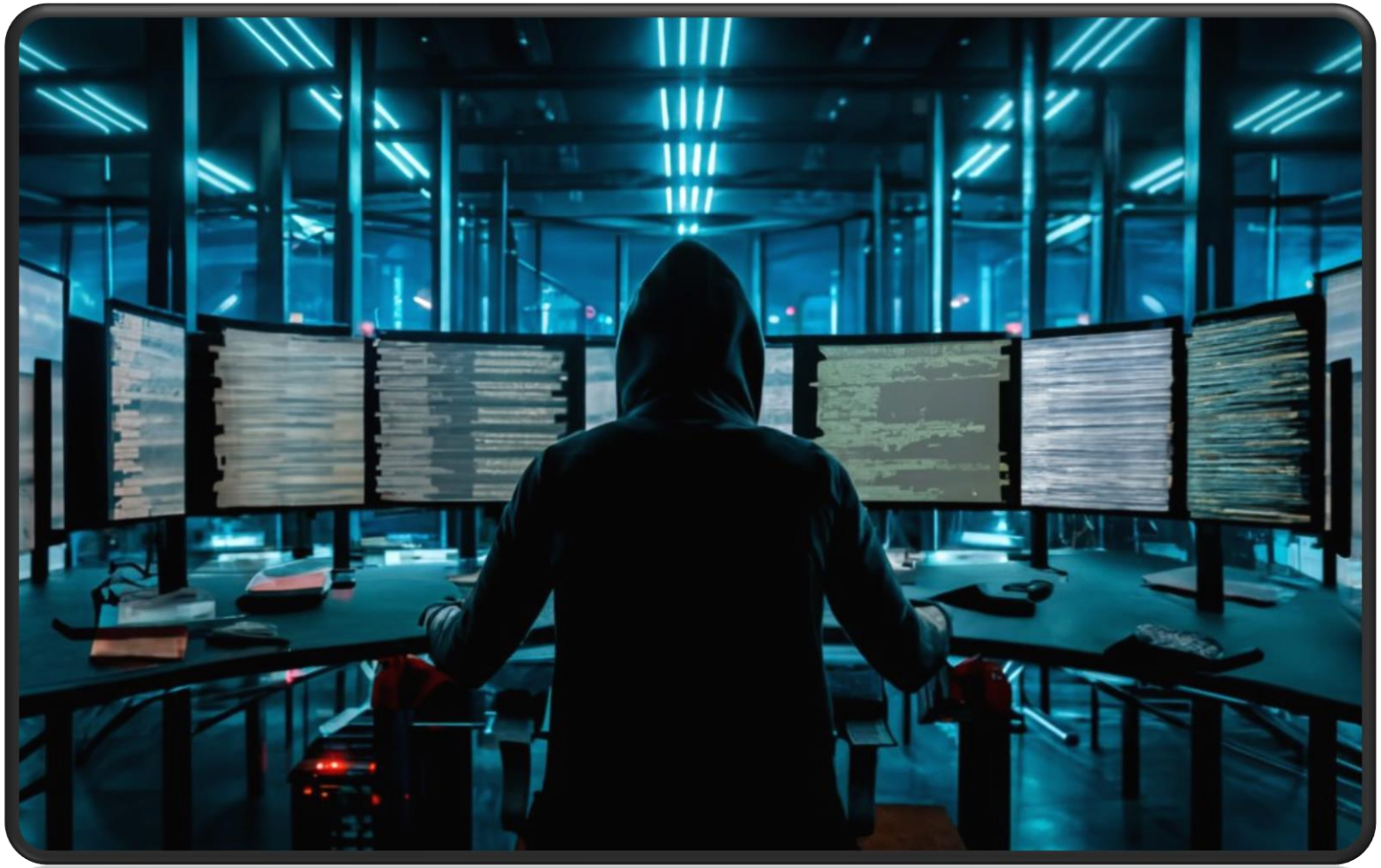


Наиболее актуальные современные угрозы

Таргетированные атаки (APT) – направлены на конкретные организации или государственные структуры, характерны действия в скрытной манере на протяжении длительного времени, избегая обнаружения стандартными средствами защиты.

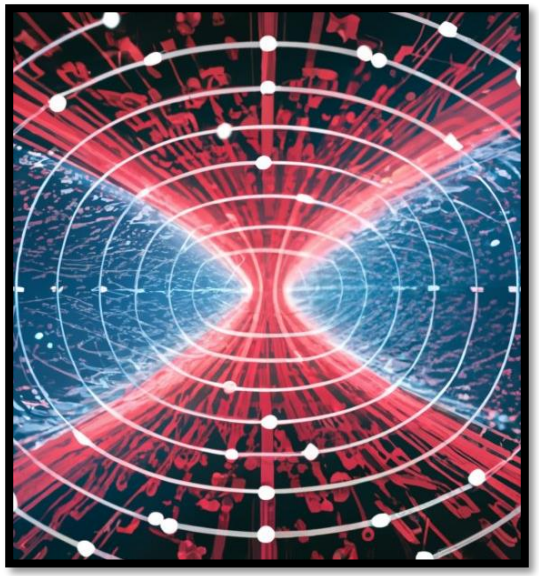
DDoS-атаки (Distributed Denial of Service) - в последнее время наблюдается рост DDoS-атак, которые направлены на выведение из строя веб-сайтов, сервисов и сетей путём перегрузки их трафиком. Атаки становятся всё более масштабными и трудно отслеживаемыми, поскольку они распределены по большому количеству источников по всему миру.

Социальная инженерия - современные атаки часто основываются на манипуляции людьми, такими как фишинг, что делает защиту намного труднее, поскольку цель — не программные уязвимости, а человеческий фактор.



Таргетированные атаки (APT)

Фаза APT	Описание	Как помогает ИИ	Классы средств с ИИ
1. Разведка	Киберпреступники собирают информацию о цели (компания, инфраструктура, сотрудники), чтобы определить уязвимости	ИИ анализирует входящий трафик и поведение пользователей, выявляя подозрительные действия. Пример: ИИ замечает, что кто-то из сотрудников получает необычно много запросов на подключение к его аккаунтам	Системы анализа сетевого трафика (NTA), инструменты Threat Intelligence
2. Начальная компрометация	Атака начинается с фишинга, внедрения вредоносного ПО через уязвимости нулевого дня, или использования слабых мест	ИИ анализирует содержимое писем и активность программ, выявляя фишинг и аномалии	Системы антиспама, решения на основе sandbox-анализа
3. Укрепление в системе	Злоумышленники закрепляют своё присутствие для повторного доступа к системе	ИИ отслеживает изменения в конфигурации системы. Пример: ИИ замечает, что сервер внезапно открывает подозрительные порты и автоматически блокирует их	Системы обнаружения и предотвращения вторжений (IDS/IPS), решения Endpoint Detection and Response (EDR)
4. Избежание обнаружения	Злоумышленники скрывают свою активность и минимизируют подозрительные действия	ИИ анализирует отклонения в сетевой активности и выявляет долгосрочные скрытые атаки. Пример: ИИ замечает, что кто-то в системе «работает» по ночам, когда все в отпуске, и отправляет оповещение	Платформы безопасности для анализа поведения (User and Entity Behavior Analytics - UEBA), SIEM

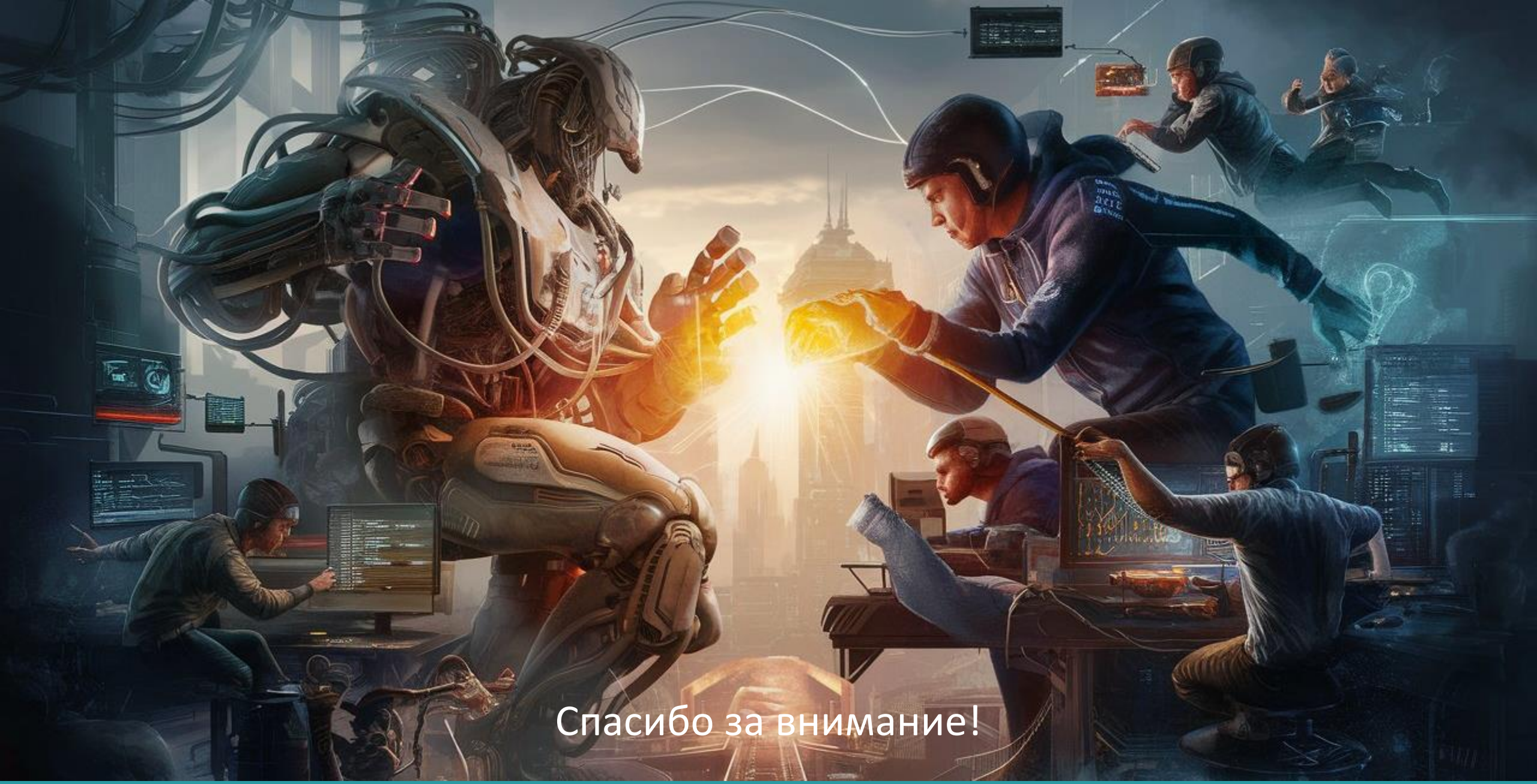


DDoS-атаки (Distributed Denial of Service)

Фаза DDoS-атаки	Описание	Как помогает ИИ	Классы средств с ИИ
1. Обнаружение аномального трафика	Атаки начинаются с массовой отправки запросов с множества источников, чтобы перегрузить систему.	ИИ отслеживает в реальном времени трафик и обнаруживает резкие всплески или нехарактерную нагрузку. Пример: ИИ замечает, что внезапно количество запросов с одного региона стало аномально большим и отмечает это как возможную атаку.	Системы анализа сетевого трафика (NTA), Web Application Firewalls (WAF)
2. Фильтрация вредоносного трафика	Вредоносный трафик направляется к серверу, чтобы блокировать доступ реальным пользователям.	ИИ использует машинное обучение для фильтрации нормального трафика от вредоносного. Пример: ИИ пропускает только «нормальные» запросы пользователей, блокируя все подозрительные запросы от атакующих.	Системы защиты от DDoS (DDoS Protection), WAF, решения Cloud Security с автоматической фильтрацией

Социальная инженерия

Фаза социальной инженерии	Описание	Как помогает ИИ	Классы средств с ИИ
1. Фишинг и обман	Злоумышленники рассылают поддельные письма, чтобы жертва выдала конфиденциальные данные или кликнула на вредоносную ссылку.	ИИ анализирует письма и выявляет фишинг на основе шаблонов и ключевых признаков	Системы защиты электронной почты с ИИ (Email Threat Protection)
2. Манипуляция людьми	Хакеры могут использовать поддельные профили в социальных сетях или притворяться коллегами, чтобы завоевать доверие.	ИИ анализирует поведение и содержание сообщений, выявляя подозрительные запросы	Платформы анализа поведения (User Behavior Analytics), Fake Detection
3. Сбор данных через фальшивые сайты	Злоумышленники создают поддельные сайты, которые имитируют настоящие, чтобы собрать логины и пароли жертв.	ИИ анализирует подозрительные сайты и предупреждает пользователей о фишинговых сайтах	Решения для фильтрации и анализа веб-контента (Web Threat Protection)



Константин Саматов