

Роль ИИ в кибербезопасности

Используя искусственный интеллект и машинное обучение, мы помогаем людям и организациям быстрее распознавать киберугрозы, снижать риски и уверенно пользоваться цифровыми технологиями в повседневной жизни и работе.

Почему это важно

Злоумышленники все активнее используют ИИ для автоматизации атак. Наши эксперты регулярно фиксируют его применение при [генерации](#) фишинговых страниц, программ-вымогателей, вредоносного ПО для продвинутых целевых атак (например, [Bluenoroff](#)), а также в кампаниях, нацеленных на российские организации, таких как [Librarian Likho](#).

Злоумышленники автоматизируют все больше этапов в цепочке атаки. Однако важно отметить, что ИИ не меняет ландшафт угроз радикально.

ИИ также активно используется в киберзащите: позволяет значительно повысить эффективность обнаружения (детектирования) угроз, включая сложные техники атак, такие как DLL Hijacking¹.

При этом надежная защита по-прежнему строится на многослойном подходе — защите конечных точек и сетей, управляемых сервисах (например, MDR), продвинутых комплексных решениях, таких как XDR, и качественной аналитике угроз (Threat Intelligence). Роль ИИ заключается в другом: он повышает скорость реакции, масштабируемость и точность защитных технологий, помогая противостоять кибератакам.

¹ Техника атаки, при которой злоумышленник внедряет вредоносную динамическую библиотеку (DLL), из-за чего легитимная программа загружает и запускает чужой код.

Как мы используем ИИ в киберзащите

~20 лет

Компания применяет технологии ИИ и ML

135 патентов

в области ИИ в портфеле интеллектуальной собственности «Лаборатории Касперского»

«Лаборатория Касперского» уже почти 20 лет [применяет](#) технологии ИИ и машинного обучения (ML) в своих продуктах и сервисах.

Искусственный интеллект помогает ежедневно анализировать сотни тысяч подозрительных и вредоносных файлов, выявляя закономерности и аномалии за доли секунды. При этом мы рассматриваем ИИ не как замену специалистам, а как инструмент поддержки: алгоритмы берут на себя рутинную обработку сигналов, освобождая экспертов для анализа сложных, целевых и нестандартных атак.

Расширяем возможности детектирования угроз

«Лаборатория Касперского» разработала множество ИИ/ML-технологий детектирования угроз, в первую очередь для выявления вредоносного ПО, а также нелегитимной активности злоумышленников. Например, в 2025 году наши эксперты [обучили](#) ML-модель выявлять попытки использования техники DLL Hijacking и усилили ей SIEM-систему KUMA. В наших решениях используется не один универсальный алгоритм, а набор специализированных моделей, каждая из которых решает свою задачу. Такой подход делает защиту более устойчивой и точной.

Ключевые направления применения ИИ/ML²

- Ранняя проверка файлов. Глубокие нейросети помогают выявлять вредоносные исполняемые файлы по статическим признакам на ранних стадиях, еще до запуска.
- Автоматизация создания правил обнаружения. ML-технологии на базе решающих деревьев помогают формировать правила обнаружения угроз, которые могут работать прямо

Наш центр экспертизы [Kaspersky AI Technology Research](#) объединяет исследователей данных, ML-инженеров, экспертов по угрозам и инфраструктуре, чтобы решать самые амбициозные задачи на стыке сфер ИИ/ML и кибербезопасности. Среди этих задач — как разработка и развитие прикладных технологий, так и проведение исследований по безопасности ИИ-алгоритмов, повышение осведомленности о рисках ИИ и многое другое.

на устройстве пользователя. Это важно, когда нужно быстро внедрить знания об угрозах в практическую защиту.

- Поведенческий анализ. Даже если файл выглядит безопасным, вредоносная активность может проявиться во время работы программы. Поведенческие модели помогают выявлять такие угрозы по нетипичным действиям.
- Выявление вредоносных интернет-ресурсов на основе анонимной телеметрии, поступающей из установленных у клиентов решений и других источников.
- Защита от фишинга и спама. Специализированные модели, включая ML-модель для детектирования мошеннических веб-страниц и DeepQuarantine для карантина писем с подозрением на спам, снижают риски пользователей.

Благодаря облачной инфраструктуре результаты работы ИИ становятся доступны пользователям почти мгновенно — новые угрозы блокируются сразу после обнаружения.

² Подробнее об этих технологиях читайте в документе [Машинное обучение для выявления вредоносного ПО](#).

Боремся с фишингом и онлайн-мошенничеством

Онлайн-мошенничество с годами становится хитрее: современные фишинговые сайты выглядят аккуратно, тексты часто написаны без ошибок, а визуальные элементы копируют интерфейсы известных сервисов. Чтобы противостоять таким угрозам, мы применяем машинное обучение и анализ контента.

В частности, для защиты от фишинга Компания использует:

- технологии оптического распознавания символов (OCR), которые выявляют вредоносный текст, спрятанный внутри изображений;
- собственные запатентованные ML-модели, обученные на массивах легитимных и поддельных сайтов и выявляющие характерные признаки мошенничества.

Это особенно важно, поскольку мошенники все чаще используют изображения и визуальные приманки, пытаясь обойти простые фильтры.

Помогаем специалистам SOC работать эффективнее

В корпоративной среде специалисты по безопасности часто сталкиваются с проблемой избыточного «шума» — большого количества уведомлений, не приводящих к реальным инцидентам. Это приводит к существенным потерям времени.

В сервисах управляемой защиты (Managed Detection and Response) ИИ-алгоритмы автоматически анализируют потоки событий и отфильтровывают ложные срабатывания, позволяя ежегодно закрывать десятки и сотни тысяч неопасных инцидентов без участия человека. В результате специалисты центров мониторинга кибербезопасности (SOC) могут сосредоточиться на действительно важных атаках и быстрее реагировать на реальные угрозы.

В корпоративных решениях для мониторинга и реагирования (Kaspersky SIEM (KUMA) и Kaspersky XDR) риск-скоринг на базе машинного обучения используется для оценки поведения устройств и серверов внутри инфраструктуры. Это помогает выявлять скрытые атаки и аномалии, не передавая данные за пределы компании, что особенно важно для организаций с высокими требованиями к конфиденциальности.

Развиваем ИИ в промышленности и на физических объектах

Искусственный интеллект используется не только для защиты компьютеров и сетей. В промышленности сбои и отклонения могут приводить к простоям, авариям и серьезным финансовым потерям.

Для таких сценариев применяются решения на базе машинного обучения, такие как [Kaspersky MLAD](#) (Machine Learning for Anomaly Detection) — программный продукт для предиктивной аналитики. Они анализируют телеметрию оборудования и помогают выявлять ранние (скрытые) признаки надвигающегося отказа оборудования, нарушения техпроцесса, кибератаки, а также ошибки персонала. Постоянно обучая нейросеть, MLAD анализирует поток «атомарных» событий от объекта, структурирует его в паттерны и выявляет нештатное поведение.

Исследуем большие языковые модели

Генеративный ИИ и большие языковые модели уже стали частью цифровой реальности. Мы развиваем инфраструктуру для исследования и безопасного использования их возможностей и быстрого создания прототипов. В этой среде развернуты LLM-инструменты, подобные ChatGPT: они доступны сотрудникам всех подразделений для решения повседневных задач и одновременно служат базой для разработки новых решений.

Одним из ключевых практических сценариев является использование языковых моделей для помощи аналитикам. В частности, сервис [Kaspersky Threat](#)

Предотвращаем злоупотребление ИИ злоумышленниками

Атакующие активно используют искусственный интеллект для автоматизации своей работы: создания фишинговых ресурсов, ускорения создания вредоносного кода, масштабирования мошеннических схем и создания аудио- и видеодипфейков.

Изучая кейсы обнаруженных угроз, созданных с применением ИИ, наши специалисты улучшают эффективность защиты от вредоносных программ, фишинга и скама, делятся рекомендациями с пользователями. Например, они советуют критически относиться к неожиданным сообщениям, аудио и видео, всегда перепроверять информацию и использовать надежные защитные решения.

[Lookup](#) (один из сервисов портала [Kaspersky Threat Intelligence Portal](#)) в 2025 году получил ИИ-инструмент для работы с OSINT-данными¹: алгоритмы анализируют открытые источники и формируют краткую сводку по индикаторам компрометации из всех публикаций, связанных с интересующим объектом. Это позволяет специалистам по информационной безопасности быстрее получать контекст и принимать решения без необходимости вручную изучать большое количество материалов.

¹ Open Source Intelligence — инструменты и возможности для работы с открытыми источниками информации.

Примеры обнаруженных зловредов, созданных с применением ИИ или использующих тему ИИ

- **FunkSec.** Анализ программы-шифровальщика показал признаки автоматической генерации фрагментов кода; среди целей — организации госсектора, IT, финансов и образования в Европе и Азии.
- **RevengeHotels.** Новая волна атак на отели с целью кражи данных банковских карт; в кампании выявлены образцы, созданные с применением ИИ.
- Зловреды под видом **DeepSeek и Grok.** Кампании с поддельными страницами, через которые распространялись стилер, вредоносный PowerShell-скрипт¹ и бэкдор; ссылки на один из зловредных ресурсов размещались в том числе в соцсети X (бывший Twitter).
- **BrowserVenom.** Фишинговый ресурс, имитирующий сайт DeepSeek, предлагал скачать модель для Windows, а на деле распространял троянец, перехватывающий трафик.
- **Jarka.** Вредоносные пакеты, распространявшиеся через репозиторий Python Package Index (PyPI) под видом инструментов для чат-ботов на основе нейросетей, заражали устройства стилером.
- **Gipy.** Загрузчик распространялся под видом приложения на основе нейросетей для изменения голоса.

¹ Скрипт для среды PowerShell, применяемый злоумышленниками для скрытого выполнения команд, загрузки вредоносного кода или управления зараженной системой.

² Способ манипулирования ИИ, при котором инструкции для модели скрыто встраиваются в данные, с которыми она работает, а не передаются ей напрямую.

Наши исследования по теме ИИ

Как нейросети выдают мошенников:

мы проанализировали артефакты, которые LLM могут оставлять на фишинговых и скам-страницах (включая характерные фразы и следы в разметке).

Для чего и как люди манипулируют нейросетями:

выяснили, как и для чего люди используют не прямые инъекции затравки², — например, чтобы обратить внимание больших языковых моделей на свои резюме и т. д.

Услуги по созданию видео- и аудиодипфейков в режиме реального времени:

проанализировали предложения на теневых площадках и типовые сценарии злоупотреблений.



Как мы управляем безопасностью ИИ

Мы рассматриваем безопасность ИИ как комплексную задачу, которая выходит далеко за рамки самих технологий. Речь идет не только о защите алгоритмов и моделей, но и о том, как ИИ используется внутри Компании, какие данные с ним работают, какие риски это создает и как построено управление этими рисками.

Вопросы безопасности ИИ затрагивают сразу несколько направлений:

- юридические и правовые аспекты — например, какие данные допустимо использовать и передавать в облачные ИИ-сервисы;
- IT- и ИБ-процессы, в том числе управление доступами, контроль конфигураций и предотвращение использования так называемого «теневого ИИ», когда сотрудники применяют внешние ИИ-инструменты без согласования;
- процессы разработки и обучения моделей, где важно понимать источники данных, возможные искажения, уязвимости моделей и сценарии их некорректного использования.

Для эффективного управления этими рисками требуется участие специалистов из разных областей — информационной безопасности, IT, юридических команд, а также экспертов по данным (data science) и машинному обучению. В идеальной модели необходимы отдельные специалисты по безопасности ИИ, координирующие работу этих команд.

Мы также учитываем, что область безопасности ИИ развивается очень быстро. Постоянно появляются новые подходы, уязвимости, методы атак и инструменты защиты. Поэтому мы рассматриваем управление рисками ИИ как непрерывный процесс, который требует регулярного обновления знаний, практик и обучающих [материалов](#).

Развиваем сотрудничество и делимся экспертизой

Экспертиза «Лаборатории Касперского» в области ИИ формируется системно — от фундаментальных исследований и инфраструктуры обучения моделей до практического применения в продуктах и сервисах. Мы исходим из того, что устойчивый прогресс в развитии ИИ возможен только при активном обмене знаниями. Поэтому Компания участвует в международных инициативах и отраслевых альянсах, сотрудничает с профессиональным и академическим сообществом и присутствует в профильных рейтингах.

В 2024 году «Лаборатория Касперского» стала участником Глобального альянса по искусственному интеллекту для промышленности и производства ([AIM Global](#)), созданного в 2023 году. Альянс объединяет правительства, международные организации, коммерческие компании и отраслевых лидеров. Участие в AIM Global позволяет обмениваться экспертизой, участвовать в формировании единых подходов к применению ИИ и поддерживать развитие технологий с учетом этических, социальных и технологических аспектов.

В 2025 году Компания присоединилась к числу организаций, поддерживающих концепцию [Глобального цифрового договора ООН](#). В документе перечислены задачи и принципы, которые будут способствовать достижению инклюзивного, открытого и устойчивого цифрового будущего.

Мы делимся с сообществами своей экспертизой в сфере ИИ. Некоторые исследования, например о [монотонных алгоритмах машинного обучения](#) или [применении нейросетей к детектированию спама](#), выходят

в формате академических статей на ведущих конференциях по машинному обучению, другие — на специализированных порталах и ИБ-конференциях.

Так, мы публикуем исследования безопасности собственных ИИ-алгоритмов — в том числе пишем о смоделированных атаках на алгоритмы [детектирования спама](#) и [вредоносного ПО](#). Исследуем использование нейронных сетей для [анализа временных рядов](#).

Кроме того, мы выпускаем обучающие материалы и курсы, в том числе для специалистов по информационной безопасности и разработчиков, чтобы помочь им безопасно внедрять ИИ-решения и учитывать возможные риски.

В 2025 году «Лаборатория Касперского» выпустила [курс](#) по обучению разработчиков и специалистов по информационной безопасности основам безопасности систем на базе больших языковых моделей. Также мы постоянно обновляем портфолио [тренингов для экспертов](#), чтобы учебные материалы соответствовали потребностям бизнеса, государственных и научных учреждений.



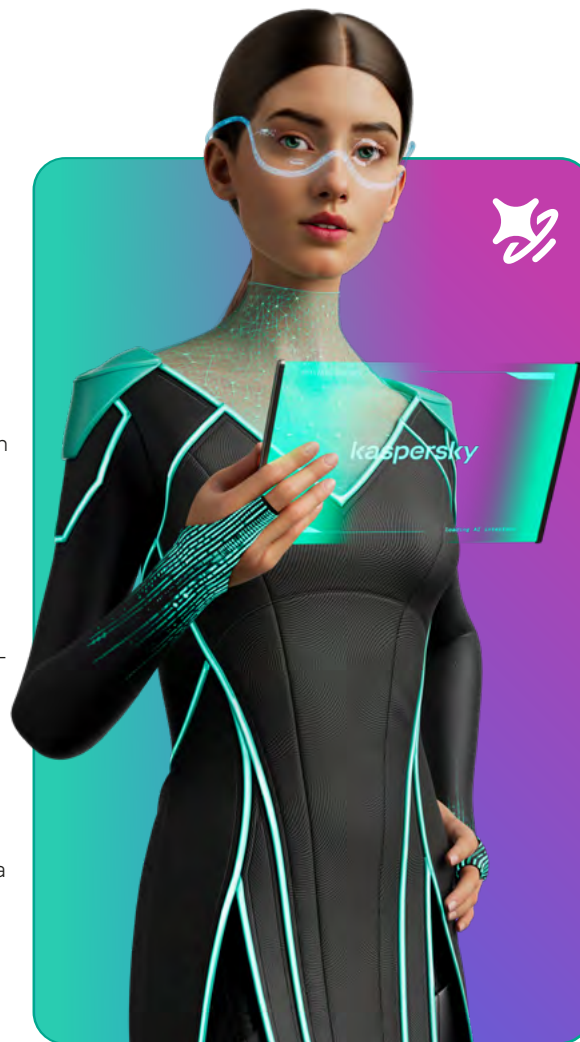
Kaspersky Unified Monitoring and Analysis Platform

Экспертиза «Лаборатории Касперского» в области ИИ и ML подтверждается и внешним признанием. В 2025 году Компания стала победителем премии «ИТ-Лидер» в номинации «Искусственный интеллект». Награды была удостоена SIEM-система [Kaspersky Unified Monitoring and Analysis Platform \(KUMA\)](#) со встроенным ИИ-ассистентом Kaspersky Investigation and Response Assistant (KIRA).

[ИИ-ассистент KIRA](#) был внедрен в платформу в конце 2024 года. В его основе лежит нейросетевая модель «ГигаЧат», разработанная Сбером. Ассистент помогает анализировать события информационной безопасности, снижая нагрузку на специалистов, сокращая объем рутинных операций и уменьшая вероятность ошибок при интерпретации инцидентов.

Smart Ranking

Также в 2025 году «Лаборатория Касперского» заняла пятое место в российском рейтинге [SmartRanking](#) в области ИИ, продемонстрировав заметный рост выручки от продаж решений с использованием ИИ и машинного обучения. А в 2024 году компания была признана одним из ключевых работодателей в сфере ИИ в России по результатам [исследования TAdviser](#).



Следуем принципам ответственного и этичного использования ИИ

Мы убеждены, что искусственный интеллект должен использоваться ответственно и прозрачно, особенно в такой чувствительной сфере, как кибербезопасность. Поэтому Компания регулярно участвует в разработке законодательных актов, создании политик и других документов, охватывающих разные аспекты безопасности в работе с новыми технологиями, в том числе ИИ.

В рамках Форума ООН по управлению интернетом в 2024 году «Лаборатория Касперского» [представила руководство](#) по безопасной разработке и внедрению систем на основе ИИ.

Ранее, в 2023 году, мы сформулировали и публично представили на Internet Governance Forum первые [принципы этичного использования ИИ](#) в кибербезопасности, которых придерживаемся в своей работе.

- **Прозрачность.** Объясняем, как работает ИИ — клиенты имеют право понимать, где и зачем используются технологии машинного обучения.
- **Безопасность.** Делаем безопасность приоритетом — все ИИ-системы проходят проверки, тестирование и специальный аудит. Принимаются меры по минимизации зависимости от сторонних наборов данных в процессе обучения ИИ-решений.
- **Человеческий контроль.** Специалисты всегда могут вмешаться, проверить и скорректировать работу алгоритмов при анализе сложных угроз.
- **Конфиденциальность.** Принимаем меры для защиты данных и систем, чтобы обеспечить цифровую приватность клиентов.
- **Приверженность целям кибербезопасности.** Сосредоточившись исключительно на защитных технологиях, мы следуем своей миссии строить более безопасный мир и демонстрируем приверженность защите пользователей и их данных.
- **Открытость к диалогу.** Обмениваемся передовым опытом в области этичного использования алгоритмов машинного обучения со всеми заинтересованными сторонами.

Планы на 2026–2027 годы

Компания планирует и дальше развивать портфель решений на базе машинного обучения, дополняя его технологиями, которые охватывают полный цикл работы с решениями по кибербезопасности. Наш подход основан на принципе разумной достаточности: мы используем максимально надежные и быстрые методы ИИ, но при этом — эффективные и понятные.

В числе приоритетных планов:

- развитие классических статистических и ML-моделей для детектирования разных видов вредоносного ПО, контентных атак и аномалий — в том числе для поиска атак, связанных с боковым перемещением и несанкционированным использованием учетных записей;
- применение генеративного ИИ для новых классов задач, недоступных традиционным ML-алгоритмам, в первую очередь через расширение навыков ассистента KIRA и применение его в разных продуктах — от Kaspersky SIEM до Kaspersky Container Security;
- развитие агентной парадигмы: создание навыков KIRA, которые комбинируют существующие функции решений и помогают в многошаговых сценариях расследования и реагирования без заранее заданного сценария. Такая глубокая интеграция может потребовать создания интерфейсов MCP (Model Context Protocol), для организации доступа ИИ-модели к встроенным инструментам наших решений;
- выход на рынок решений по управлению уязвимостями (Vulnerability Management). Новый продукт, усиленный технологиями ИИ, поможет организациям своевременно выявлять и устранять уязвимости и ошибки конфигурации в IT-инфраструктуре.