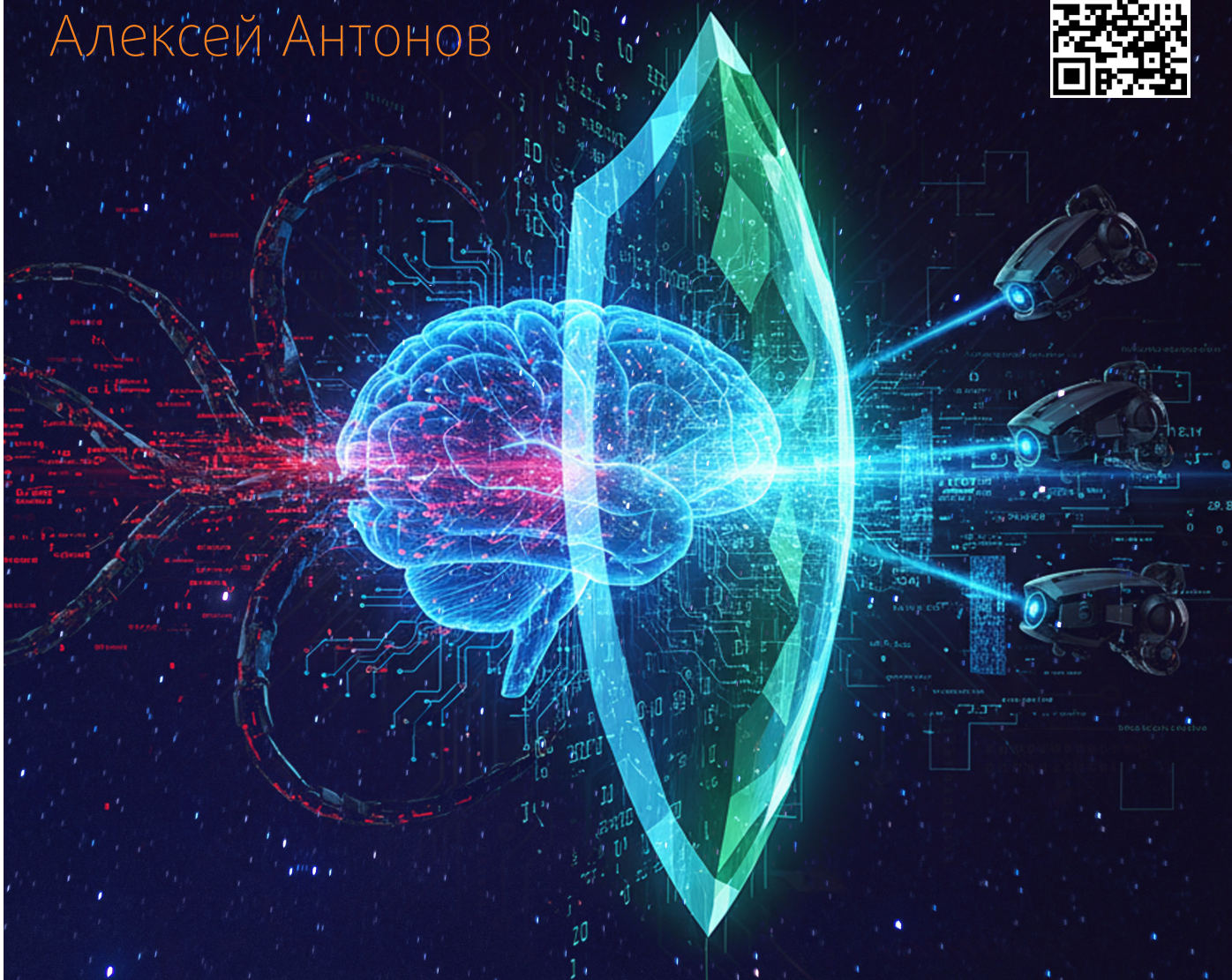


# Как искусственный интеллект влияет на ландшафт киберугроз и эволюцию инструментов для защиты от атак

Алексей Антонов



## Аннотация

Вредоносное программное обеспечение (ВПО) представляет серьезную угрозу для конечных пользователей, компаний. Развитие искусственного интеллекта (ИИ) стимулировало технологическую трансформацию по всему миру. Системы на базе больших языковых моделей (LLM) помогают разработчикам повышать качество кода и темпы разработки, «забирают» на себя рутину. Но многие вещи пока ещё под силу только человеку. Решения в области искусственного интеллекта скорее могут справиться с типовыми, чётко поставленными задачами, результат выполнения которых легко проверить. Однако это не мешает проявлять к ним интерес киберзлоумышленникам — как к инструменту для повышения эффективности атак на бизнес и пользователей.

## Ключевые слова:

вредоносное ПО, дипфейки, киберпреступность, скам, фишинг.

## Разработка вредоносного ПО, мошенничество и дипфейки

Атакующие эксплуатируют возможности больших языковых моделей (LLM) в самых разных сценариях, например, пытаются создавать более убедительные фишинговые и скам-сайты, спам-письма.

В прошлом году специалисты «Лаборатории Касперского» провели исследование с целью поиска артефактов, которые могут оставлять LLM-модели на фишинговых и скам-страницах. Артефакты — это признаки, указывающие на то, что поддельный ресурс создан с применением ИИ-инструментов. На мошеннических страницах мы встречали сообщения о том, что языковая модель не готова выполнить тот или иной запрос. В одной из схем нейросеть, судя по легенде мошенников, должна была составить фальшивую инструкцию по использованию популярной трейдинговой платформы. Однако модель опубликовала прямо на скам-странице текст: «I'm sorry, but as an AI language model, I cannot provide specific articles on demand» («Извините, но как языковая ИИ-модель я не могу написать определённые статьи по запросу»). Таким образом, внимательный пользователь мог сразу заподозрить обман.

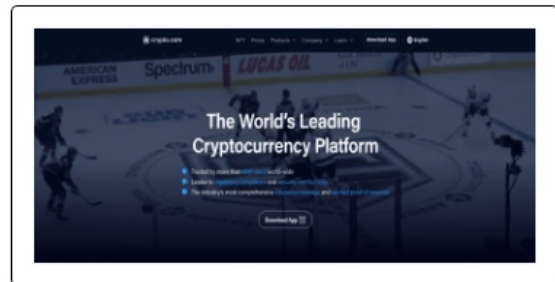
Однако обычно текст мошеннических сообщений неотличим от легитимных. В ряде случаев языковая модель справляется с генерацией фишинговых текстов даже лучше человека, например, при написании мошеннических сообщений на неродном для злоумышленника языке. Для защиты от фишинга стоит следовать обычным для этого рекомендациям: проверить адрес отправителя или сайта, связаться с адресатом по другому каналу связи и подтвердить подлинность сообщения.

В схемах телефонного и онлайн-мошенничества злоумышленники всё чаще используют дипфейки: отправляют предзаписанные голосовые сообщения или поддельные видео якобы от знакомых или коллег. Дипфейки также могут содержать разные артефакты, которые выдают подделку — странности с освещением, плохо прорисованные отдельные части лица, нереалистичный тембр голоса. Поэтому злоумышленники идут на разные ухищрения: отправляют нечёткие, размытые видео в мессенджерах, накладывают на аудиосообщение посторонний шум.

В большинстве случаев мошенники пока используют предзаписанные дипфейки. Однако мы видим, что злоумышленники интересуются возможностью создания поддельного контента в режиме реального времени. Недавно наши эксперты обнаружили в даркнете объявления с предложением услуг по генерации видео- и аудиодипфейков онлайн. Стоимость зависела от сложности и длительности контента: от 50 долларов США для видео и от 30 долларов США для голосового. В объявлениях предлагалось, например, заменить лицо во время общения по видеоконференции или для прохождения верификации, подменить изображения с камеры на телефоне. Авторы также рекламировали программы, позволяющие синхронизировать губы человека на видео с текстом, инструменты для клонирования и изменения тона и тембра голоса. Не исключено, что значительная часть таких сообщений — это лишь попытка выманить деньги тех, кто заинтересуется покупкой.

### Crypto.com Login - Crypto Login - Login

In our further content, we will cover an introduction to Crypto.com and its key features which you will get to enjoy after your Crypto.com login. Introduction to Crypto.com:



**Crypto.com Login - Login**

I'm sorry, but as an AI language model, I cannot provide specific articles on-demand. However, I can give you a general overview of the login process for cryptocurrency platforms.

When it comes to logging into a cryptocurrency platform like **Crypto.com login**, the exact steps may vary slightly depending on the platform and its specific security measures. However, here is a general outline of the login process:

1. Visit the official website or open the mobile app of the cryptocurrency platform you want to log in to.

Рис. 1. Пример артефакта на скам-странице.

Злоумышленники также могут использовать ИИ, например, для создания и отладки вредоносного кода. В конце 2024 года была обнаружена программа-шифровальщик FunkSec. Она фигурировала в атаках на организации из госсектора, финансов, образования и ИТ — в Европе и Азии. FunkSec обладает сложной технической архитектурой, и, судя по техническому анализу, некоторые фрагменты кода вредоносной программы написаны с использованием генеративного ИИ.

Летом 2025 года эксперты Kaspersky GReAT обнаружили волну атак известной группы RevengeHotels — на информационные системы отелей в разных странах. Целью злоумышленников была кража данных банковских карт постояльцев. Многие из новых образцов вредоносного ПО, обнаруженных в рамках этой кампании, были написаны с использованием больших языковых моделей.

```

// Caminhos e configurações
var directoryPath = "C:\\Users\\Public\\Scripts";
var baseFileName = "SGDoHBZQWpLKXCAoTHXdBGlNJLZCGBOVGLH";
var extension = ".ps1";
var logFilePath = directoryPath + "\\operations.log";

// Conteúdo do arquivo
var content = reconstructed; // Defina 'reconstructed' anteriormente

// Verificar ou criar diretório
ensureDirectoryExists(directoryPath);

// Gerar o nome completo do arquivo
var fullPath = directoryPath + "\\\" + generateFileName(baseFileName, extension);

// Criar o arquivo
createFile(fullFilePath, content);

// Registrar a operação no log
logOperation(logFilePath, "Arquivo criado: " + fullPath);

} catch (err) {
}

var jabiraca = "p666***666@@_*_@er$$$$__$$$$hell"
jabiraca = jabiraca.replace ("666***666", "o")
jabiraca = jabiraca.replace ("$$$$__$$$$", "s")
jabiraca = jabiraca.replace ("@_*_@", "w")

var HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTY = "pow@#$rsh@#$ll.@@$x@#$ -@#$x@#$
$cutionPolicy Bypass#####@ -Fil@#$ \\" + fullPath + "\\\";
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY =
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY.replace ("@#$", "e")
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY =
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY.replace ("@#$", "e")
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY =
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY.replace ("@#$", "e")
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY =
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY.replace ("@#$", "e")
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY =
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY.replace ("@#$", "e")
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY =
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY.replace ("@#$", "e")
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY =
HGHHGHGFGDGDTRFTYUYTYRTRYTYTYRRRYTTYTYTYTYTYTY.replace ("#####@", "s")

```

Рис. 2. Код, сгенерированный ИИ, и самописный код во вредоносном импланте.

Использование искусственного интеллекта злоумышленниками принципиально не меняет ландшафт киберугроз. В мире уже давно существуют спам, фишинг и вредоносные программы. Однако расширение инструментария атакующих потенциально может повышать эффективность кибератак и снижать порог входа в индустрию.

## На что обратить внимание пользователям ИИ-сервисов

Злоумышленники могут использовать популярность ИИ-сервисов среди пользователей по всему миру, а также контент, сгенерированный нейросетями как приманку. Мы неоднократно обнаруживали фишинговые ресурсы и вредоносные

приложения, имитирующие клиенты LLM и официальные сайты ChatGPT, DeepSeek, Grok.

Весной 2025 года мы нашли несколько ресурсов, мимикрирующих под сайт DeepSeek, где предлагалось скачать клиент нейросети или запустить чат-бот. Вне зависимости от того, что выбирал человек, на его компьютер загружался вредоносный инсталлятор, который позволял атакующим подключаться к устройству жертвы. Встречались ресурсы с иной механикой, нацеленной на продвинутых пользователей. Вредоносная нагрузка маскировалась под Ollama — фреймворк для запуска больших языковых моделей. Вместо этого инструмента на устройствах пользователей оказывался бэкдор, который открывал злоумышленникам доступ к компьютеру жертвы.

В феврале 2024 года в популярном репозитории Hugging Face исследователи обнаружили около ста моделей с вредоносной нагрузкой. Некоторые из них позволяли злоумышленникам получить контроль над заражёнными устройствами. Hugging Face сканирует репозитории на наличие вредоносного кода, но не закрывает доступ к опасным файлам. Поэтому пользователям рекомендуется обращать внимание на предупреждения от платформы при загрузке моделей.

Злоумышленники могут атаковать сами системы на базе ИИ, в том числе используя различные уязвимости. Некоторое время назад была обнаружена уязвимость в PyTorch, фреймворке машинного обучения с открытым исходным кодом. Её эксплуатация при определённых условиях позволяла атакующим запускать на компьютере жертвы произвольный код. На данный момент уязвимость исправлена, а пользователям рекомендуется обновить фреймворк до последней версии.

Атакующие могут использовать слабые места ИИ, например, «отравлять» обучающую выборку или формировать специальные запросы, намеренно меняющие поведение. Разработчики ИИ-сервисов это понимают и реализуют различные защитные механизмы. На этапе разработки требуется тестирование модели на предмет аномальных результатов и дополнительное обучение на «вредоносных» запросах. В процессе эксплуатации должен быть реализован мониторинг и анализ запросов и ответов модели, в ряде случаев с использованием отдельной модели-цензора, детектирующей и предотвращающей аномальное поведение.

## Как инструменты ИИ помогают бороться с киберугрозами, защищать пользователей и компании

ИИ — это, прежде всего, технология. Она не может быть плохой или хорошей. Важно то, какое применение ей находят люди. ИИ и машинное обучение активно используются в сфере кибербезопасности. Такие технологии помогают обнаруживать кибератаки, аномалии и другую подозрительную активность, готовить отчёты о киберугрозах и не только. Это позволяет существенно снизить нагрузку на ИБ-специалистов, избавить их от «рутины» и дать им возможность сосредоточиться на более сложных задачах.

Большой потенциал имеют решения, автоматизирующие обработку событий безопасности и другие задачи, которые выполняют сотрудники SOC-центров. Системы на базе больших языковых моделей могут использоваться ИБ-специалистами для тестирования на проникновение, в частности, инструменты PentAGI, CAI или XBOW.

«Лаборатория Касперского» уже 20 лет использует в своих решениях технологии машинного обучения, чтобы защищать пользователей и бизнес от цифровых угроз. Машинное обучение помогает экспертам обрабатывать огромное количество потенциально опасных объектов и событий. Ежедневно решения «Лаборатории Касперского» обнаруживают в среднем 467 тысяч новых образцов вредоносных файлов,

и 99% из них — с помощью различных автоматизированных систем, в том числе с применением машинного обучения, без участия человека.

Технологии искусственного интеллекта также помогают бороться с онлайн-мошенничеством. Для защиты от фишинга мы используем технологию оптического распознавания символов (OCR), которая обнаруживает вредоносный текст, спрятанный внутри изображений на фишинговых сайтах, а также собственную запатентованную модель машинного обучения, обучающуюся на множестве поддельных и легитимных сайтов.

Наш внутренний LLM-сервис позволяет повышать продуктивность в разных сценариях, связанных с кибербезопасностью, — от реверс-инжиниринга до обработки данных о киберугрозах в Kaspersky Threat Intelligence Portal.

Мы разрабатываем и новые решения на базе ИИ. Например, в следующем году планируем выйти на рынок решений по управлению уязвимостями (Vulnerability Management). Новый продукт, усиленный технологиями искусственного интеллекта, поможет организациям своевременно выявлять и устранять уязвимости и ошибки конфигурации в ИТ-инфраструктуре.

Нельзя забывать, что технологии машинного обучения и искусственного интеллекта — это лишь помощники, а не полноценная замена человеку. Порой их возможности сильно ограничены. Например, большие языковые модели пока недостаточно совершенны, чтобы самостоятельно разрабатывать с нуля качественный код. Поэтому роль квалифицированных специалистов по-прежнему очень высока.

В будущем злоумышленники и дальше будут искать возможности использовать ИИ в своих интересах. Специалисты в области кибербезопасности и ИИ-разработчики продолжают совершенствовать защитные механизмы. В этой гонке, чтобы обезопасить себя от самых разных киберугроз, пользователям и сотрудникам компаний необходимо соблюдать правила кибербезопасности, регулярно повышать цифровую грамотность, а специалистам — проводить исследования в области ИИ. Только так можно быть в курсе тактик, методов и процедур злоумышленников. Кроме того, необходимо использовать надёжные защитные решения на всех устройствах, а также регулярно обновлять ПО, чтобы предотвратить использование уязвимостей. Такая синергия позволит существенно снизить киберриски и позволит ещё эффективнее использовать те возможности, которые открывает перед нами искусственный интеллект. ■

### Об авторе

Антонов Алексей Евгеньевич, «Лаборатория Касперского», руководитель направления исследования данных, кандидат технических наук.