SMART Закупка

29 мая 2023

Киберпреступность встречается с ChatGPT: берегись, мир

Мир гудит от того, на что способен ChatGPT . Конечно, он отвечает как на мирские, так и на философские вопросы , пишет код и отлаживает его и даже может помочь в выявлении болезни Альцгеймера . Но, как и в случае с любой новой технологией, чат-бот OpenAI на базе ИИ рискует быть использованным не по назначению.

Исследователи из Check Point Software обнаружили, что ChatGPT можно использовать для создания фишинговых писем. В сочетании с Codex , системой преобразования естественного языка в код, также разработанной OpenAI, ChatGPT можно использовать для разработки и внедрения вредоносного кода . «Наши исследователи построили полную цепочку заражения вредоносным ПО, начиная с фишингового письма и заканчивая документом Excel, содержащим вредоносный код VBA [Visual Basic for Application]. Мы можем скомпилировать все вредоносное ПО в исполняемый файл и запустить его на машине», — говорит Сергей Шикевич , руководитель группы анализа угроз в Check Point Software. Он добавляет, что ChatGPT в основном создает «гораздо более качественные и убедительные фишинговые электронные письма и электронные письма, выдающие себя за другое лицо, чем настоящие фишинговые электронные письма, которые мы видим сейчас в дикой природе».

ChatGPT «позволит большему количеству людей стать кодерами, но самый большой риск заключается в том, что больше людей могут стать разработчиками вредоносных программ». —Сергей Шикевич, Check Point Software

Тем не менее, повторение является ключевым моментом, когда речь идет о ChatGPT. «Что касается кода, то первый вывод не был идеальным, — говорит Шикевич. «Я бы сравнил то, как я использую его, с Google Translate, где результат в основном будет хорошим. Но я проверю это и внесу некоторые исправления или корректировки. То же самое происходит с ChatGPT, где вы не можете использовать код в том виде, в котором он есть, и необходимо внести небольшие коррективы».

Лорри Фейт Крэнор , директор и почетный профессор Bosch Института безопасности и конфиденциальности CyLab и профессор компьютерных наук, инженерии и государственной политики FORE Systems в Университете Карнеги-Меллона, разделяет это мнение. «Я не пробовал использовать ChatGPT для генерации кода, но я видел несколько примеров от тех, кто это делал. Он генерирует код, который не так уж и сложен, но некоторые его части на самом деле работоспособны», — говорит она. «Существуют и другие инструменты искусственного интеллекта для генерации кода, и все они становятся лучше с каждым днем. ChatGPT, вероятно, сейчас лучше генерирует текст для людей и может особенно хорошо подходить для создания таких вещей, как реалистичные поддельные электронные письма».

Исследователи также выявили хакеров, использующих ChatGPT для разработки вредоносных инструментов , таких как похититель информации и рынок даркнета. «[ChatGPT] позволит

большему количеству людей стать программистами, но самый большой риск заключается в том, что больше людей могут стать разработчиками вредоносных программ», — говорит Шикевич.

«Я думаю, что для успешного использования этих инструментов [ИИ] сегодня требуются некоторые технические знания, но я ожидаю, что со временем станет проще использовать результаты этих инструментов и запускать атаку», — говорит Крэнор. «Поэтому, хотя неясно, что то, что инструменты могут сделать сегодня, вызывает гораздо больше беспокойства, чем инструменты, разработанные человеком, которые широко распространены в Интернете, не пройдет много времени, прежде чем эти инструменты будут разрабатывать более сложные атаки, с возможностью быстро генерировать большое количество вариантов».

Дальнейшие сложности могут возникнуть из-за отсутствия способов определить, был ли вредоносный код создан с помощью ChatGPT. «Нет хорошего способа точно определить, что конкретное программное обеспечение, вредоносное ПО или даже фишинговое письмо было написано ChatGPT, потому что у него нет подписи», — говорит Шикевич.

Со своей стороны, OpenAI работает над методом нанесения «водяных знаков» на результаты моделей GPT , которые впоследствии можно будет использовать для доказательства того, что они были созданы ИИ, а не людьми. Шикевич также отмечает, что после того, как Check Point Software опубликовала свои выводы, исследователи обнаружили, что больше невозможно генерировать фишинговые электронные письма с помощью ChatGPT.

Чтобы защититься от этих угроз, создаваемых искусственным интеллектом, Шикевич советует компаниям и частным лицам принять соответствующие меры кибербезопасности. Текущие меры безопасности по-прежнему применяются, и очень важно продолжать обновлять и улучшать эти реализации.

«Исследователи также работают над способами использования ИИ для обнаружения уязвимостей кода и обнаружения атак», — говорит Кранор. «Надеюсь, наступление обороняющейся стороны сможет не отставать от наступления атакующей стороны, но это еще предстоит выяснить».

Хотя системы с поддержкой ИИ, такие как ChatGPT, обладают огромным потенциалом для изменения того, как люди взаимодействуют с технологиями, они также несут риски, особенно при опасном использовании.

«ChatGPT — отличная технология, которая может демократизировать ИИ, — говорит Шикевич. «ИИ был чем-то вроде шумной функции, которую понимали только специалисты по информатике или алгоритмам. Теперь люди, не разбирающиеся в технологиях, начинают понимать, что такое ИИ, и пытаются внедрить его в свою повседневную жизнь. Но самый большой вопрос в том, как бы вы его использовали и для каких целей?»

Ссылка на статью: Киберпреступность встречается с ChatGPT: берегись, мир