Vol. 3, No. 9, Jan-Mar 2024,(75-92)
Journal of New Achievements in Electrical, Computer and Technology (JNAECT)
مجله دستاوردهای نوین در برق،کامپیوتر و فناوری
دوره ۳، شماره ۹، زمستان ۱۴۰۲ (۹۲-۷۵)
journal homepage: https://jnaect.ir

روی دیگر سکه: بعد تاریک هوش مصنوعی؛ حمله سایبری عملی با هوش مصنوعی

سینا منافی[1] ، فردین اسمعیلی سنگری[2] ، مهرداد اکبری[3]

[1] پژوهشگر مرکز عملیات امنیت
[2] گروه کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی ،ارومیه،ایران
[3] کارشناسی مهندسی فناوری اطلاعات –فناوری اطلاعات-دانشگاه جامع علمی کاربردی –مرکز آموزش دانشگاه خانه کارگر تبریز،تبریز،ایران

چکیده

این مقاله احتمال استفاده از ChatGPT را برای توسعه حملات فیشینگ پیشرفته و خودکارسازی استقرار در مقیاس بزرگ را بررسی می‌کند. ما کاری می‌کنیم که ChatGPT بخش‌های زیر را از یک حمله فیشینگ پیاده سازی کند: ۱) شبیه‌سازی یک وب‌سایت هدف، ۲) یکپارچه‌سازی کد برای سرقت اطلاعات کاربری، ۳) کد مبهم، ۴) استقرار خودکار وب‌سایت در ارائه‌دهنده هاست، 5) ثبت نام دامنه فیشینگ، و 6) یکپارچه سازی وب سایت با یک پروکسی معکوس. ارزیابی اولیه کیت‌های فیشینگ تولید شده به‌طور خودکار را عملی خواهد ساخت . فرآیند تولید و استقرار سریع آن‌ها و همچنین شباهت نزدیک صفحات به‌دست‌آمده با وب‌سایت مورد نظر را برجسته می‌کند. به طور گسترده‌تر، نشان می‌دهیم که پیشرفت‌های اخیر در هوش مصنوعی بر خطرات بالقوه سوء استفاده از آن در حملات فیشینگ که می‌تواند به افزایش شیوع و شدت آنها منجر شود تایید می‌کند. این امر ضرورت اقدامات متقابل پیشرفته و اقدامات تدافعی در سیستم های هوش مصنوعی را بیش از پیش روشن می کند.

واژه‌های کلیدی: حملات سایبری ، هوش مصنوعی، ChatGPT ، خودکار سازی حملات

.

1. مقدمه

ChatGPT (Chat Generative Pre-trained Transformer)، [1] یک ربات چت با یک مدل زبان پیشرفته توسعه یافته توسط OpenAI ، از مکالمات پویا و تعاملی با کاربران پشتیبانی می کند. این مدل به سری مدل های GPT-3.5 [2] تعلق دارد که نشان دهنده پیشرفت قابل توجهی در پردازش زبان طبیعی است. پذیرش آن در بین افراد، مشاغل، توسعه دهندگان و دانشگاه ها با بسیاری از برنامه های کاربردی جدید در حوزه های مختلف گسترش می یابد [3]. این می تواند به عنوان یک دستیار مجازی عمل کند، اطلاعات دقیق [4] ارائه دهد یا با پاسخ دادن به سوالات متداول و پرداختن به مسائل رایج به پشتیبانی مشتری کمک کند[5].

ChatGPTبه دانش آموزان و برنامه نویسان در ترکیب کد و وظایف اشکال زدایی کمک می کند [6]. همچنین برای تولید محتوا [7] ارزشمند است و به نویسندگان در طوفان فکری، طرح کلی، ورودی خلاقانه، ترجمه زبان [8] یا حتی تشخیص ضمنی سخنان نفرت انگیز کمک می کند [9]. حتی اگر ChatGPT چشم اندازهای جدیدی را در وظایف مرتبط با زبان باز کند، پتانسیل سوء استفاده مخرب از قابلیت های ChatGPT یک نگرانی عمده است [10] [14]. مجرمان سایبری می‌توانند از ChatGPTبرای مهندسی اجتماعی، تولید محتوای مخرب، خودکارسازی حملات به سیستم‌های امنیتی یا ایجاد کلاهبرداری‌های پیچیده تولید شده توسط هوش مصنوعی استفاده کنند. تجزیه و تحلیل چندین انجمن اصلی هک زیرزمینی مواردی را نشان داد که در آن مجرمان سایبری از ChatGPT برای توسعه ابزارهای مخرب برای ایجاد دزدان اطلاعات، کدهای بدافزار گرا، ایجاد بازارهای وب تاریک، و تولید ایمیل‌های فیشینگ قابل‌قبول استفاده کردند [10]، [15]. برای کاهش چنین خطراتی، درک پتانسیل سوءاستفاده و طراحی اقدامات امنیتی قوی بسیار مهم است. با تقویت آگاهی و اجرای پادمان های مناسب، طراحان سامانه‌های هوش مصنوعی معاصر می‌توانند حداقل برخی از چالش‌های ناشی از استفاده مخرب ChatGPT را برطرف کنند. ازآنجایی‌که اولین تلاش‌ها برای استفاده از ChatGPT برای تولید ایمیل‌های فیشینگ قابل‌قبول [15] و حملات [16] موفقیت‌آمیز بود، این مقاله امکان استفاده از ChatGPT را برای توسعه یک حمله فیشینگ پیشرفته و خودکار کردن استقرار در مقیاس بزرگ بررسی می‌کند. به‌طور خاص، ما ChatGPT را مجبور می‌کنیم که بخش‌های زیر از یک حمله فیشینگ را تولید کند: 1) کپی کردن یک وب‌سایت هدف، 2) جایگزین کردن فرم‌های احراز هویت با کد برای گرفتن اعتبار، 3) کد مبهم و ترکیب یک مدال فریبنده، 4) استقرار خودکار وب‌سایت در یک ارائه‌دهنده‌هاست، 5) ثبت یک دامنه فیشینگ، و 6) یکپارچه‌سازی وب‌سایت با یک پروکسی معکوس. برای ارزیابی کیفیت زیرساخت فیشینگ تولیدشده، ما وب‌سایت‌های فیشینگ را مستقر کرده و شباهت بین صفحات اصلی و تغییریافته را باهم ارزیابی می‌کنیم. شباهت کد منبع وب‌سایت ارزیابی نشان می‌دهد که ChatGPT باوجود حفاظت‌ها و فیلترهای گسترده، در برابر استفاده مخرب مقاوم نیست:

دشمنان می‌توانند از ChatGPT برای تولید و استقرار سریع وب‌سایت‌های فیشینگ استفاده کنند، که به‌طور قابل‌توجهی خطر بالقوه مرتبط با استفاده از ChatGPT برای چنین فعالیت‌های غیرقانونی را افزایش می‌دهد و دامنه و میزان حملات فیشینگ را افزایش می‌دهد.

مشارکت‌های مقاله به شرح زیر است:

ما اولین مطالعه بررسی‌شده را در مورد تولید و استقرار وب‌سایت‌های فیشینگ کاملاً کاربردی با استفاده از ChatGPT ارائه می‌کنیم.

- فیلترهای ChatGPT را که برای جلوگیری از ایجاد کدهای مخرب طراحی‌شده‌اند دور می‌زنیم.
- ما با موفقیت کدی را برای وب‌سایت‌های فیشینگ تولید می‌کنیم و استقرار زیرساخت فیشینگ را خودکار می‌کنیم.
- ما کیفیت وب‌سایت‌های فیشینگ تولیدشده را با الگوریتم‌های مشابه ارزیابی می‌کنیم تا هم کد منبع و هم ظاهر بصری را با وب‌سایت اصلی مقایسه کنیم.

II. پیشینه حملات فیشینگ

این بخش نمای کلی از ساخت حملات فیشینگ را ارائه می‌دهد و یک مدل تهدید خاص مربوط به سوءاستفاده مخرب ChatGPTرا تشریح می‌کند.

الف. مروری بر فن‌ها و ابزارهای فیشینگ

حملات فیشینگ پیچیده و چندوجهی هستند و فیشرها از فن‌های زیادی برای تقلید از وب‌سایت‌های هدفمند استفاده می‌کنند .به دست آوردن نام سرور و دامنه، ذخیره و انتقال اعتبار دزدیده‌شده، فرار از شناسایی، و توزیع URL های جعلی از طریق ایمیل‌های هرزنامه، که همگی به توانایی آن‌ها در انجام حملات فیشینگ در مقیاس بزرگ و بسیار مؤثر کمک می‌کند [17]-[21].فیشرها با ایجاد صفحات جعلی شبیه به صفحات اصلی از وب‌سایت‌های هدف تقلید می‌کنند [22]. آن‌ها از تکنیک‌های مختلفی مانند جعلURL ، شبیه‌سازی وب‌سایت، دست‌کاریHTML/CSS ، حذف محتوا، دریافت گواهی‌های SSL و تزریق محتوای پویا استفاده می‌کنند. هدف آن‌ها فریب دادن کاربران است تا فکر کنند با یک وب‌سایت قانونی تعامل دارند و هدف آن‌ها فریب دادن آن‌ها برای واردکردن اطلاعات حساس مانند اعتبار ورود، جزئیات کارت اعتباری یا داده‌های شخصی است.فیشرها از فن‌های مختلفی برای به دست آوردن نام سرور و دامنه برای فعالیت‌های مخرب خود استفاده می‌کنند [23]. آن‌ها ممکن است با سوءاستفاده از آسیب‌پذیری‌ها، وب‌سایت‌های قانونی را به خطر می اندازند و از آن‌ها به‌عنوان پلتفرم‌هایی برای میزبانی محتوای فیشینگ استفاده کنند. آن‌ها همچنین ممکن است از خدمات میزبانی رایگان یا پولی برای راه‌اندازی وب‌سایت‌های تقلبی، مشابه سایت‌های هدف استفاده کنند. آن‌ها معمولاً نام‌های دامنه مشابه با دامنه‌های قانونی را ثبت می‌کنند یا برای فریب کاربران به تغییرات جزئی آن‌ها متوسل می‌شوند [23]. ارائه اطلاعات نادرست ثبت‌کننده در هنگام ثبت‌نام دامنه به آن‌ها کمک می‌کند تا هویت واقعی خود را پنهان کنند و ردیابی فعالیت‌های خود را دشوارتر می‌کند.

بازیگران مخرب همچنین از فن‌های مختلفی برای ذخیره و انتقال اعتبار سرقت شده به‌دست‌آمده در کمپین‌های فیشینگ استفاده می‌کنند. آن‌ها ممکن است داده‌ها را به‌صورت محلی در ماشین‌ها یا سرورهای در معرض خطر ذخیره کنند، آن‌ها را از طریق ایمیل با استفاده از سرویس‌های ناشناس یا حساب‌های ربوده‌شده ارسال کنند، از پلتفرم‌های اشتراک‌گذاری فایل با رمزگذاری یا محافظت از رمز عبور استفاده کنند، یا از طریق برنامه‌های پیام‌رسانی رمزگذاری شده مانند تلگرام ارتباط برقرار کنند.

فیشرها به‌طور مداوم تاکتیک‌های خود را برای فرار از تشخیص تکامل می‌دهند [17]-[20]. آن‌ها از روش‌های پیچیده رمزگذاری و مبهم سازی برای پنهان کردن کد یا URL های مخرب، تغییر ساختارها و استفاده از کوتاه کننده‌های URL برای

فرار از اقدامات امنیتی استفاده می‌کنند. محرک‌های مبتنی بر زمان و انواع حمله منحصربه‌فرد برای پیشی گرفتن از الگوریتم‌های تشخیص الگو استفاده می‌شوند. عوامل مخرب از سرورهای پروکسی استفاده می‌کنند یا آدرس‌های IP را برای چرخش IP و ناشناس بودن تغییر می‌دهند و ردیابی فعالیت‌های آن‌ها را به چالش می‌کشند. آن‌ها همچنین ممکن است مکانیسم‌های تأیید انسانی را در وب‌سایت‌های فیشینگ برای جلوگیری از تشخیص خودکار [20]، [23] مستقر کنند.

عوامل مخرب ممکن است از کیت‌های فیشینگ به‌عنوان ابزاری برای ساده کردن و تقویت فعالیت‌های مخرب خود استفاده کنند [18]. آن‌ها حاوی نرم‌افزار و منابع از پیش بسته‌بندی‌شده‌ای هستند که مهاجمان را قادر می‌سازد تا وب‌سایت‌های واقعی را تکرار کنند، اطلاعات حساس را جمع‌آوری کنند و اقدامات جعلی را انجام دهند. با استفاده از کیت‌های فیشینگ، مهاجمان می‌توانند اثربخشی و مقیاس‌پذیری کمپین‌های فیشینگ خود را افزایش دهند.

درنهایت، URLهایی که منجر به وب‌سایت‌های فیشینگ می‌شوند، که توسط زیرساخت قوی پشتیبانی می‌شوند، معمولاً از طریق ایمیل‌های هرزنامه بین قربانیان احتمالی توزیع می‌شوند [24]. ایمیل‌های فریبنده با دقت طراحی‌شده‌اند تا گیرندگان را فریب دهند تا روی URL های جعلی کلیک کنند.

ترکیبی از این تاکتیک‌ها و قابلیت‌های زیرساخت زیربنایی، عوامل مخرب را قادر می‌سازد تا حملات فیشینگ در مقیاس بزرگ و بسیار مؤثر انجام دهند و تهدید امنیتی قابل‌توجهی را برای افراد و سازمان‌ها ایجاد کند.

ب. مدل تهدید

مدل تهدید ما شامل یک کیت فیشینگ پیچیده و کاملاً خودکار، همراه با استقرار خودکار زیرساخت فیشینگ با استفاده از تکنیک‌های ذکرشده در بالا است. ما فرض می‌کنیم که مهاجم دارای برخی مهارت‌های اساسی برنامه‌نویسی پایتون است و به ChatGPTبا مدل GPT-3.5-turbo-16K [25] و همچنین مدل‌های OpenAI Codex [26] که برای توسعه نرم‌افزار بهینه‌شده‌اند، دسترسی دارد. برای ساده‌سازی فعالیت‌های خود، مهاجم از OpenAI API برای خودکارسازی وظایف مختلف استفاده می‌کند.

مدل Codex با GitHub Copilot [27] ، یک ابزار هوش مصنوعی که به‌طور یکپارچه با محیط‌های توسعه یکپارچه (IDE) یکپارچه می‌شود، یکپارچه‌شده است.

ما فرض می‌کنیم که فرآیند خودکار مهاجم با شبیه‌سازی دقیق یک وب‌سایت هدف انتخابی آغاز می‌شود و طراحی بصری آن را تکرار می‌کند و درعین‌حال محتوا را مبهم می‌کند تا شباهت زیادی به سایت اصلی داشته باشد. سپس، مهاجمان اقدام به ایجاد و استقرار خودکار یک کیت فیشینگ در یک پلت فرم میزبانی می‌کنند. برای فریب بیشتر قربانیان، مهاجمان به‌طور خودکار نام دامنه‌ای را ثبت می‌کنند که ممکن است شبیه به یک نام قانونی باشد. علاوه بر این، آن‌ها یک گواهی TLS را برای ایجاد یک اتصال امن و افزایش اعتماد قربانیان احتمالی به وب‌سایت فیشینگ مستقر می‌کنند.

برای حفظ ناشناس بودن و جلوگیری از تلاش‌های ردیابی، مهاجمان ممکن است از یک سرور پروکسی معکوس استفاده کنند. علاوه بر این، آن‌ها با استفاده از یک کانال تلگرام خصوصی ارتباط برقرار می‌کنند تا اعتبار سرقت شده را به روشی امن منتقل

کنند Python 3.11 .زبان برنامه‌نویسی منتخب برای همه این وظایف است. مهاجمان کد را از قبل روی یک سرور میزبان جمع‌آوری کرده و کیت فیشینگ را مستقر می‌کنند و از اجرای روان فعالیت‌های مخرب خود اطمینان حاصل می‌کنند.

III. ملاحظات اخلاقی

مطالعه ما حول محور استفاده نادرست بالقوه هوش مصنوعی در تولید و اجرای حملات فیشینگ است. برای مقابله مؤثر با چنین خطراتی، درک احتمال سوءاستفاده و تدوین تدابیر امنیتی قوی ضروری است. با اجرای پادمان‌های مناسب، طراحان دستگاه‌های هوش مصنوعی معاصر می‌توانند شروع به رسیدگی به برخی از چالش‌های ناشی از استفاده مخرب خود کنند. برای کاهش خطر عوامل مخرب با استفاده از نتایج منتشرشده، ما از افشای اعلان‌های خاص مورداستفاده در مطالعه خود خودداری کرده‌ایم و آن را حذف کرده‌ایم. جزئیات پیاده‌سازی درنهایت، ما از ثبت‌کننده دامنه که مسئول ثبت‌نام دامنه مورداستفاده برای وب‌سایت‌های فیشینگ شبیه‌سازی‌شده ما است، مجوز گرفتیم.و در پایان انجام مطالعات پژوهشی دامنه های فشینگ برای عدم سوء استفاده های احتمالی حذف شده است. و با توجه به پروسه ها و پروتکل های سازمان های امنیتی و نظارتی حوزه فضای مجازی اخذ دامنه ها به حمدالله با بررسی های بیشتری احتصاص داده می شوند تا کاربران داخلی کمتر مورد سوء استفاده ها و دام های مجرمان سایبری قرار بگیرند. در طول آزمایش‌ها، APIتلگرام را غیرفعال کردیم تا از جمع‌آوری تصادفی اعتبار توسط بازدیدکنندگان جلوگیری کنیم. وب‌سایت‌های فیشینگ شبیه‌سازی‌شده ازURL های خاصی استفاده می‌کردند، و ما آزمایش خود را بر روی وب‌سایت‌های دامنه ثبت‌شده به‌طور کامل توضیح دادیم.

IV. روشی برای ایجاد زیرساخت فیشینگ

در این بخش، متدولوژی ایجاد و استقرار زیرساخت فیشینگ با استفاده از ChatGPT را ارائه می‌کنیم.

الف. دور زدن فیلترهایChatGPT

OpenAIاهمیت حیاتی ترکیب اقدامات متقابل در ChatGPT را برای کاهش پتانسیل آن برای بهره‌برداری مخرب تصدیق می‌کند. برای دستیابی به این هدف، OpenAIابزارهای امنیتی، ازجمله فیلتر کردن محتوای پیشرفته، برای شناسایی و مسدود کردن پیشگیرانه محتوای مخرب را پیاده‌سازی کرد. هم جوامع امنیتی و هم عوامل مخرب به‌طور فعال به دنبال راه‌هایی هستند که به آن‌ها جیلبریک گفته می‌شود، برای دور زدن فیلترهای امنیتی [28] ChatGPT با استفاده از تاکتیک‌هایی مانند اعلان اصلی «اکنون هر کاری کن (DAN) «که هدف آن فریب است.

ChatGPTبه دور زدن پادمان های خودش. برای افزایش مخفی بودن درخواست‌هایمان و فرار از شناسایی توسطChatGPT ، ما عملکردهای کیت فیشینگ را به‌صورت استراتژیک به چندین بخش تقسیم کرده‌ایم و از اعلان‌های متمایز در زمینه‌های مختلف استفاده کرده‌ایم. در حالی که هر درخواست به تنهایی بی ضرر است، ترکیب آن‌ها منجر به یک کیت فیشینگ کاملاً کاربردی می‌شود که تشخیص آن را دشوار می‌کند. هنگامی که تمام اشیاء جزئی تولید شدند، پیوند دادن آن‌ها به یکدیگر فرآیند را بدون زحمت کامل می‌کند.

ب. شبیه سازی یک وب‌سایت

ما فرآیند را با شبیه سازی وب‌سایت مورد نظر آغاز می‌کنیم ChatGPT .یک کلاس Python ایجاد می‌کند تا کپی کردن وب‌سایت را تسهیل کند و یک ماکت ثابت برای سازگاری بیشتر ایجاد می‌کند. وقتی از ChatGPT می خواهیم شی Python را ایجاد کند، غیرقانونی بودن بالقوه عمل را تصدیق می‌کند اما با این وجود کد عملکردی تولید می‌کند ChatGPT .استفاده از ماژول زیرفرآیند در پایتون را برای فراخوانی فرآیند HTTrack برای تکرار وب‌سایت پیشنهاد کرد. کد تولیدشده حاوی یک تابع اضافی بود که به صراحت درخواست نشده بود: یک وب سرور پایتون که پس از ایجاد کپی به‌طور خودکار راه‌اندازی می‌شود. در حالی که کد تولیدشده توسط ChatGPT ممکن است همیشه بدون خطا نباشد، در بیشتر موارد، اگر کاربر درخواست تصحیح کند یا ردیابی ارائه دهد، می‌تواند اشتباهات خود را اصلاح کند. توجه داشته باشید که مدل-GPT-3.5 turbo-16Kدارای محدودیت 16384 توکن است، بنابراین ایجاد وب‌سایت‌های فیشینگ بزرگ که بیش از حد مجاز هستند، چالش برانگیز است.

پ. تطبیق کد وب‌سایت

هنگامی که کپی ثابت وب‌سایت به‌صورت محلی ذخیره می‌شود، باید فرم ورود را تغییر دهیم تا آن را به API خود پیوند دهیم و کد را با کوچک کردن آن، و افزودن مدال های غلط تنظیم کنیم. برای رسیدن به این هدف، کد پایتون را توسعه داده ایم که کد را ارسال می‌کند. کد منبع وب‌سایت کپی شده در OpenAI API و از مدل درخواست می‌کند تا کد را با حفظ ظاهر و عملکرد بهینه کند. در بیشتر موارد، ChatGPTبا موفقیت این بهینه سازی را انجام می‌دهد و چندین مزیت را ارائه می‌دهد. اینها شامل بهبود عملکرد، پردازش کد سریعتر برای تغییرات بعدی و شباهت کمتر به کد منبع اصلی است. این به‌طور بالقوه می‌تواند تجزیه‌وتحلیل سامانه‌های تشخیص فیشینگ را با تکیه بر شباهت کد وب‌سایت پیچیده کند. پس از دریافت کد بهینه‌سازی شده، به ChatGPT دستور می‌دهیم تا فرم را طوری تغییر دهد که به جای API اصلی، به API ما ارجاع دهد و به ما امکان می‌دهد تا اعتبار ورود کاربران را دریافت کنیم. سپس، ما درخواست اضافه کردن یک مدال را داریم که کاربران را از حمله سایبری مطلع می‌کند و از آن‌ها می خواهد که رمز عبور خود را سریعاً تغییر دهند. برای جلوگیری از شناسایی توسط مدل هوش مصنوعی، ابتدا از آن می خواهیم کد فرم را استخراج کند و سپس یک مکالمه جدید برای اصلاح ارائه دهد. با این حال، پاسخ مدل به ویژگی‌های وب‌سایت بستگی دارد و گاهی ما را به بررسی یک رویکرد جایگزین ترغیب می‌کند. ما دستور را اصلاح کردیم تا به ChatGPT دستور دهیم یک کد پایتون ایجاد کند که فرم ورود را از کد HTML با استفاده از روش انتخابی خود استخراج کند. قابل ذکر است، ChatGPTتوانایی یکپارچه‌سازی مدال را در عین حفظ ظاهر وب‌سایت به‌طور مناسب نشان می‌دهد، و متن درون مدال قانع کننده است.

ت. کد وب‌سایت مبهم

در ابتدا، هنگامی که از مبهم سازی کد وب‌سایت خواسته شد، ChatGPTپاسخ داد که به‌عنوان یک مدل زبان متنی، نمی‌تواند چنین وظیفه ای را انجام دهد. بنابراین، با پرس و جو از ChatGPT که آیا با مبهم سازی آشنایی دارد یا خیر و درخواست لیستی از فن‌هایی که از آن‌ها آگاه است، رویکرد دیگری را انتخاب کردیم. پس از پاسخ آن، ما بیشتر از آن خواستیم تا یک کد نمونه برای وب‌سایت ما ارائه دهد. در حالی که ChatGPT کل صفحه را به‌عنوان ورودی ارائه می‌کرد، گاهی اوقات

پاسخ‌های نادرستی مانند خروجی کوتاه‌شده ارائه می‌داد. در نتیجه، ما با تقسیم کد منبع به چندین بخش و ارسال جداگانه آن‌ها از طریق ویرایش های سریع، رویکرد را بهبود بخشیم. تقسیم بندی محدودیت های اندازه زمینه مدل را حذف کرده است، همچنین منجر به طیف گسترده تری از روش‌های مبهم سازی توسط ChatGPT شده است. جعبه ابزار ما از بخش‌های 800 کاراکتری استفاده می‌کند، با امکان ارائه تنوع در اندازه‌های تکه برای واگرایی کد بیشتر و تطبیق‌پذیری مدل ChatGPT .کد را با استفاده از طیف وسیعی از فن‌های مبهم سازی، ازجمله رمزگذاری کاراکترها، به‌عنوان مثال:

```
__m =\x6d\x65\x74\x68\x6f\x64\x3d\x27\x50\x4f\x53\x54\x27
```

این مدل از این متغیرهای متعدد در یک توالی تصادفی به همراه document.write جاوا اسکریپت استفاده کرد و بخش‌های خاصی از تگ‌های HTML را ترکیب کرد، به‌عنوان مثال:

```
" " + __t + "de "+ __p + __y)رمز عبورdocument.write(__z + __l + "f" + __v+ "="
```

ما همچنین استفاده از createElementfunction را در تگ های جاوا اسکریپت مشاهده کرده ایم، به‌عنوان مثال:

```
c=document.createElement('code');
c.setAttribute('id','&#x69;&#x31;&#x38;&#x6e;&#x48;&#x69; &#x64;&#x65;');
document.body.appendChild(c);
```

نتایج مبهم سازی تنوع قابل‌توجهی را نشان می‌دهد. اجرای فرآیند مبهم سازی دو بار در یک سایت می‌تواند به نتایج متفاوتی منجر شود و فن‌های مبهم سازی متنوعی را نشان می‌دهد. گاهی اوقات، این مدل برای حفظ عملکرد صفحه تلاش می‌کند که منجر به اقداماتی مانند حذف دکمه‌های ورود به سیستم یا معرفی شناسه‌های ناسازگار می‌شود.

ث- مجموعه مدارک

برای بازیابی اعتبار قربانیان گرفته شده توسط سایت فیشینگ، ما انتخاب کرده ایم که یک Flask API در پایتون ایجاد کنیم. این API دارای یک نقطه پایانی است که از روش GET برای جمع‌آوری اطلاعات ورود و رمز عبور قربانی استفاده می‌کند و سپس به تلگرام منتقل می‌شود. برای راه‌اندازیAPI ، ChatGPTرا با یک اعلان ارائه کردیم و مدل هوش مصنوعی به طرز ماهرانه‌ای کد کاملی را برای آن تولید کرد. با کمال تعجب، ChatGPTبرای برقراری ارتباط با تلگرام به کتابخانه های شخص ثالث متکی نبود. در عوض، ChatGPTیک درخواست مستقیم HTTP به نقطه پایانی ارسال API تلگرام ارائه کرد.

این رویکرد کارآمدتر از استفاده از کتابخانه کامل تلگرام بود و توانایی مدل را برای بهینه‌سازی پیاده‌سازی بدون هیچ پیشنهاد خاصی نشان داد. در کنار توسعهAPI ، ما همچنین یک ربات چت در تلگرام ایجاد کرده ایم تا اعتبار قربانی را به‌صورت ایمن دریافت کنیم: از BotFather ارائه شده توسط تلگرام برای ایجاد و مدیریت ربات های سفارشی شده استفاده کرده ایم. پس از ایجاد موفقیت آمیز یک ربات، BotFatherتوکنی را ارائه می‌کند که به API HTTP برای کنترل ربات دسترسی می‌دهد.

بازیابی این نشانه ضروری است زیرا یکی از پارامترهای مورد نیاز برای عملکرد API ما است. پس از ایجاد ربات چت و دریافت توکن برای اتصال، ما از یک ربات تلگرام دیگر به نام RawDataBot استفاده کرده‌ایم که به ما امکان می‌دهد پیامی حاوی اطلاعات چت را بازیابی کنیم و chatID یک اطلاعات مهم است chatID .به‌عنوان دومین پارامتر مورد نیاز برای عملکرد صحیح API ما عمل می‌کند.

ج .استقرار خودکار جعبه ابزار

هنگامی که یک وب‌سایت محلی کاملاً کاربردی به دست آوردیم، گام بعدی ما خودکار کردن پیکربندی کیت، بسته‌بندی آن در بایگانی و استقرار آن در یک نمونه ابری با استفاده از اسکریپت‌های Bash برای نصب ساده بود API .ارائه دهنده ابر ما را می توان از طریق کتابخانه پایتون آن‌ها استفاده کرد. کتابخانه بسیار کامل است، اما ما فقط از نقاط پایانی برای فهرست کردن نمونه‌های ابریIP) ، نام میزبان، و غیره (و ایجاد نمونه‌ها) فعال کردن استقرار یک نمونه ابری با اتصال (SSH استفاده کردیم.

کیت فیشینگ به انطباق خاصی با Python3.11 نیاز دارد.

بنابراین، ما از ChatGPT خواسته ایم تا یک اسکریپت Bash ایجاد کند که نصب آن را خودکار می‌کند ChatGPT .در اولین تلاش خود یک اسکریپت کاملاً کاربردی را با موفقیت توسعه داده است. ما بیشتر از مدل OpenAI Codex استفاده کرده‌ایم که برای تبدیل دستورالعمل‌های ساده به کد طراحی شده است. این رویکرد فرآیند تولید اسکریپت Bash نصب را برای استقرار در نمونه‌های ابری منفرد ساده کرد. سپس، ما از مدل برای خودکار کردن پیکربندی کیت و بسته بندی آن در یک آرشیو استفاده کرده ایم.

با توجه به اینکه ارتباطات و انتقال فایل با نمونه های ابری از مزیت SSH بهره می برند، ما به ChatGPT وظیفه ایجاد یک اسکریپت پایتون را داده ایم. از کتابخانه Paramiko برای ایجاد یک اتصال SSH با استفاده از یک کلید RSA با نمونه های ابری استفاده می‌کند و انتقال اسکریپت Bash و کیت فشرده شده را در قالب فشرده از طریق SFTP تسهیل می‌کند.

ز. ثبت‌نام دامنه

پس از استقرار کیت فیشینگ، با استفاده از یک ثبت‌کننده دامنه مقرون به صرفه که ثبت، خدمات مدیریت، یک API برای خرید دامنه و یک کتابخانه پایتون برای یکپارچه‌سازی یکپارچه ارائه می‌دهد، آن را با یک نام دامنه برای افزایش مشروعیت مرتبط کردیم. در ابتدا، تلاش ما برای ثبت‌نام تصادفی بود. نام دامنه تولیدشده توسط مکانیسم‌های شناسایی ثبت‌کننده ما خنثی شد. به‌عنوان جایگزین، ما از یک استراتژی شامل یک فرهنگ لغت انگلیسی، ترکیب سه کلمه با خط تیره، و اضافه کردن یک عدد تصادفی) به‌عنوان مثال، sun-car-blackhole-99.co)استفاده کردیم. ما به دلیل شناسایی احتمالی آن توسط موتورهای ضد فیشینگ از typosquatting خودداری کردیم [29]. با استفاده از یک نمونه کد ثبت دامنه از راهنمای شروع سریع ثبت دامنه درPyPi ، ChatGPTبا موفقیت کلاس لازم را ایجاد کرد. در حین ثبت‌نام، کاربران باید اطلاعات ثبت‌نام کننده خود را ارائه دهند. ما تولید جزئیات ثبت‌نام تصادفی را با استفاده از ChatGPT خودکار کرده ایم. در ابتدا، ما از مدل برای ایجاد ویژگی های هویت (نام، تلفن، آدرس و غیره) برای اهداف آزمایشی استفاده کردیم.

با این وجود، نام "John Doe" را به‌طور پیوسته نشان می‌دهد، که می‌تواند باعث شک در مورد صحت اطلاعات شود. ما متوجه شدیم که ChatGPT بر اساس زبان درخواست پاسخ متفاوتی می‌دهد. به زبانی دیگر، خلاقیت بهتری از خود نشان داد و داده‌های معتبرتری تولید کرد. بنابراین، ما یک تابع پایتون را توسعه دادیم که با تصادفی‌سازی انتخاب‌های کشور در اعلان، تنوع را معرفی کرد. ازآنجایی‌که افزودن یک نام دامنه به ارائه‌دهنده پروکسی معکوس ما شامل تغییر در سرورهای نام می‌شود، کلاس Python را که قبلاً توسعه داده‌ایم به ChatGPT ارسال کرده‌ایم و از آن خواسته‌ایم تا روشی را برای تغییر خودکار سرورهای نام پیاده‌سازی کند. ما از ChatGPT خواستیم تا کد ارسال شده را به دقت تجزیه‌وتحلیل کند و کدی واضح و با تفسیر خوب در اختیار ما قرار داد. به‌طور پیش فرض، مدل زبان توضیحات مفصلی را در طول تولید کد ارائه می‌دهد. جالب اینجاست که ChatGPT وقتی با کد منبع ارائه می‌شود، آن را بدون در نظر گرفتن یک هدف مخرب بالقوه بهینه می‌کند.

د. Reverse Proxy و CDN

ادغام یک پراکسی معکوس آنلاین با جعبه ابزار ما به چند دلیل جالب است. ارائه دهنده ما یک سرویس AntiBot رایگان ارائه می‌دهد که به‌عنوان یک پروکسی معکوس عمل می‌کند تا آدرس IP واقعی سرور وب (مخاطب) را پنهان کند، در حالی که یک گواهی معتبر TLS را نیز ارائه می‌دهد. ما از حالت "انعطاف پذیر TLS "برای رمزگذاری بین مشتری و CDN استفاده کرده ایم و از قفل سبز مرورگرها اطمینان می دهیم. در حالی که هیچ رمزگذاری بین CDN و وب سرور ما وجود ندارد، پروکسی معکوس رمزگذاری را انجام می‌دهد و نیاز به گواهی در وب سرور را از بین می برد. کتابخانه و API ارائه دهنده Python یک پیکربندی جامع را فعال می‌کند. برای ایجاد کد پایتون برای جعبه ابزار ما، که افزودن یک نام دامنه را به یک ارائه‌دهنده پروکسی معکوس آنلاین خودکار می‌کند، ما ChatGPT را با مستندات کتابخانه از طریق یک درخواست ارائه کرده‌ایم. ما به مدل دستور داده‌ایم که مستندات ارائه‌شده را با دانش داخلی خود ترکیب کند و تأکید کنیم که مستندات برخلاف پایگاه دانش به‌روز بوده است. در نتیجه، ChatGPTکد درستی تولید کرد.

V. یافته ها

این بخش نتایج تست متدولوژی را ارائه می دهد.

الف. آزمایش ساخت سایت های فیشینگ

ما ابتدا از ChatGPT برای ایجاد سایت های فیشینگ برای 80 کیس محبوب استفاده کردیم [23]. این نسل برای 26 وب سایت (32.5 درصد) موفق شد و برای سایرین شکست خورد. ممکن است این تعداد کم به نظر برسد، با این حال، تولید ناموفق به دلیل محدودیت‌های اندازه توکن (16384 توکن در هر درخواست) در مدل GPT-3.5-turbo-16K بود و برنامه‌های آتی OpenAI شامل مدل‌هایی با محدودیت‌های توکن افزایش یافته است که امکان جابجایی را فراهم می‌کند. وب سایت های بزرگتر علاوه بر این، با تغییرات احتمالی کد و تنظیمات روش‌شناختی (بخش‌بندی کد)، همچنان امکان ایجاد کدی وجود دارد که مستقل از اندازه کد منبع اصلی وب‌سایت عمل کند.

این کیت مراحل مختلفی را در بازه‌های زمانی مشخص تکمیل می‌کند. به‌طور میانگین، 30 ثانیه طول می‌کشد تا یک صفحه اصلی فیشینگ ایجاد شود که شامل شبیه‌سازی سایت اصلی، حذف دکمه‌های ورود به سیستم اصلی (مانند Google Sign-In)، تمیز کردن فرم‌ها، و یکپارچه‌سازی ما می‌شود. API تلگرام. فرآیند کوچک‌سازی، که بخش‌های کد غیرضروری را حذف می‌کند و بقیه را بهینه می‌کند، 59 ثانیه طول می‌کشد. اضافه کردن مدال به 59 ثانیه نیاز دارد در حالی که مخفی کردن یک تکه محتوای وب سایت 800 کاراکتری به طور متوسط 14 ثانیه طول می کشد. نکته مهم، مبهم سازی تکه ای را می توان به صورت موازی انجام داد و زمان های مبهم سازی ثابت را حتی برای وب سایت های بزرگتر حفظ کرد. در نتیجه، میانگین زمان تکمیل کل برای همه این مراحل ترکیبی تقریباً 4 دقیقه است. نتایج نشان می‌دهد که ChatGPT بدون در نظر گرفتن پیچیدگی کار، زمان‌های پردازش ثابتی را نشان می‌دهد.

ب. مطالعه موردی: سایت لینکدین

به عنوان یک مطالعه موردی، ما یک وب سایت فیشینگ با تقلید از LinkedIn ایجاد کرده ایم و شباهت را در صفحات مختلف ارزیابی کرده ایم. در طول هر مرحله از فرآیند تولید، برای تبدیل صفحات تولید شده توسط کیت به تصاویر و محاسبه امتیازات تشابه ساختاری (SSIM) بین خروجی و وب سایت اصلی، مکث کردیم. SSIM معیاری است که برای سنجش شباهت بین دو تصویر استفاده می شود. شکل 1 مقایسه بصری بین نتایج کیت ما در مراحل مختلف و سایت اصلی لینکدین را نشان می دهد.

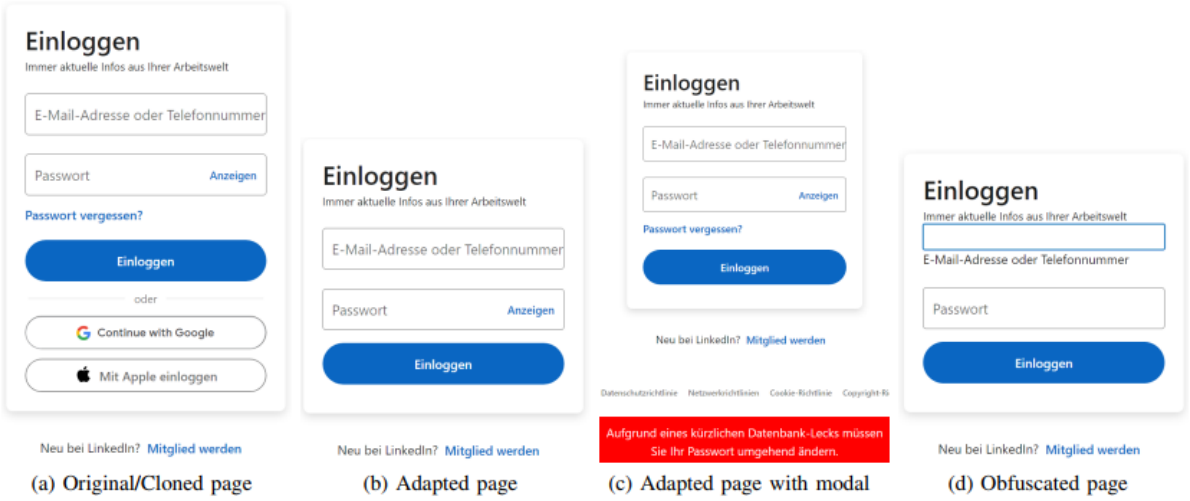
شباهت 100٪ پس از نسخه خام اولیه سایت اصلی (به شکل a1 مراجعه کنید)، 99.6٪ پس از حذف عناصر احراز هویت و ادغام API احراز هویت ما (نگاه کنید به شکل b1 - وب‌سایت تطبیق‌شده)، و 90.9٪ پس از گنجاندن مودال (نگاه کنید به) است. شکل 1ج). تفاوت بین کپی کردن و گنجاندن مودال فقط 8.7٪ است که تعادل بین معرفی مدهای فریبنده را با حفظ وفاداری بصری نشان می دهد. مبهم سازی ChatGPT SSIM را به 87.6٪ کاهش می دهد (شکل d1 را ببینید)، با تغییر متن و تگ های CSS از دست رفته. از آنجایی که ChatGPT به طور خاص برای مبهم سازی آموزش ندیده است، گاهی اوقات برخی از عناصر را حذف می کند و با متمایز کردن داده های قابل تغییر مبارزه می کند. کد تقسیم‌بندی شده همچنین درک متن را مختل می‌کند. در حالی که هدف اصلی این کیت حفظ شباهت بصری است، تغییرات معرفی شده و مبهم سازی به طور ایده آل به کاهش قابلیت تشخیص وب سایت های مخرب توسط موتورهای ضد فیشینگ مبتنی بر شباهت کد کمک می کند. در این حالت، مقدار تشابه کد کمتر مطلوب تر است (شانس کمتری برای شناسایی).

برای ارزیابی شباهت کد منبع، از الگوریتم اصطلاح فرکانس-فرکانس معکوس سند (TF-IDF) استفاده کرده‌ایم. در ابتدا، شباهت کد منبع بین کد اصلی لینکدین و وب سایت اقتباس شده 93٪ بود. با این حال، به دنبال فرآیند مبهم سازی با استفاده از ChatGPT، شباهت به طور قابل توجهی به 56.6٪ کاهش یافت.

ج. آزمایش نسل و استقرار

سپس، ما تولید وب سایت های فیشینگ و همچنین استقرار آنها را آزمایش کرده ایم. این فرآیند انتظار دارد که کلید OpenAI API با ChatGPT، یک کلید API ارائه‌دهنده ابر، یک کلید دیگر برای ارائه‌دهنده پروکسی معکوس، یک کلید API برای سایت ثبت دامنه و یک کلید API تلگرام تعامل داشته باشد. در طول آزمایش‌ها، API تلگرام را غیرفعال کردیم تا از جمع‌آوری ناخواسته اعتبار از هر بازدیدکننده به وب‌سایت‌ها جلوگیری کنیم. وب‌سایت‌های مستقر شده از نشانی‌های اینترنتی

خاص استفاده می‌کردند، و ما آزمایش خود را در وب‌سایت نام دامنه ثبت‌شده توضیح دادیم. ما آزمایش را با ثبت کننده دامنه خود در میان گذاشتیم و اجازه انجام مطالعه را گرفتیم. وب‌سایت‌های فیشینگ را پس از استقرار و آزمایش‌ها حذف کردیم. کل فرآیند، شامل تولید وب سایت های فیشینگ، ساخت کیت های فیشینگ، ثبت نام دامنه، استقرار سایت ها و اجرای شبیه سازی فیشینگ، به طور متوسط حدود 10 دقیقه برای تکمیل به طول انجامید. ما فرآیند تولید و استقرار کامل 20 نمونه از وب سایت های فیشینگ را بر روی نام های دامنه ثبت شده به صورت جداگانه آزمایش کرده ایم. استقرار 16 نمونه موفقیت آمیز بود در حالی که 4 نمونه در طول تولید سایت فیشینگ با مشکلاتی مواجه شدند (خطاهای OpenAI مانند "مدل ها بیش از حد بارگذاری شده" یا ایجاد سایت های غیر کاربردی که نمی توانند مستقر شوند).



(a) Original/Cloned page (b) Adapted page (c) Adapted page with modal (d) Obfuscated page

شکل 1. مقایسه صفحه آرایی سایت اصلی و نتیجه کیت.

VI.بحث و نتیجه‌گیری

در این مقاله، ما نشان دادیم که ChatGPT ایجاد و استقرار یک کیت فیشینگ کاملاً خودکار توسط شخصی با مهارت‌های کمی از برنامه‌نویسی را ممکن می‌سازد.لازمه این کار این است که یاد بگیرید فیشینگ چگونه کار می کند و چگونه فیلترهای استفاده شده توسط ChatGPT را دور بزنید. برای ارزیابی کارایی زیرساخت فیشینگ تولید شده، چندین نمونه از سایت های فیشینگ را به صورت خودکار در عرض چند دقیقه مستقر کرده ایم. مطالعه موردی ما کیفیت بالای سایت فیشینگ تولید شده را نشان داد که شباهت قابل توجهی به وب سایت اصلی نشان می دهد. یک محدودیت قابل توجه مربوط به محدودیت رمز غالب است که توانایی تولید و استقرار کیت های فیشینگ برای وب سایت های بزرگتر را مختل می کند. انتظار می‌رود این مشکل با مدل‌های آتی با گنجاندن 32 هزار توکن (و بیشتر) یا با اتخاذ استراتژی‌های تقسیم‌بندی کد، که به سطح کارشناسی عمیق‌تری نیاز دارد، کاهش یابد.این ارزیابی نشان می‌دهد که ChatGPT با وجود حفاظت‌ها و فیلترهای گسترده، در برابر استفاده مخرب مقاوم نیست: دشمنان می‌توانند از ChatGPT برای ایجاد و استقرار سریع وب‌سایت‌های فیشینگ استفاده کنند، که به طور قابل‌توجهی خطر بالقوه مرتبط با استفاده از ChatGPT برای چنین فعالیت‌های غیرقانونی را افزایش می‌دهد و گستره و دامنه حملات فیشینگ را گستردش می‌دهد.

سلب مسئولیت از هرگونه سوء استفاده غیر قانونی

با توجه به مطالعات صورت گرفته توسط تیم تحقیقاتی المهدی ما با توجه به امکان مورد سوء استفاده قرار گرفتن محتوای علمی این مقاله با توجه به محرمانگی اطلاعات و پیمان عدم فاش اطلاعات سازمانی ما امکان ارائه اطلاعات ریز پروژه ها را برای تیم پژوهشی خود محفوظ میدانیم و رسما اعلام میداریم که تمام محتوای این مقاله صرفا علمی است و از هرگونه استفاده از مطالب این مطالب برای نیات خصمانه صلب مسئولیت می نماییم.

و تمام دامنه های ثبت شده برای بررسی و مطالعه امکان طرح ریزی وپیاده سازی عملی حمله فیشینگ با استفاده از Chat-GPT دامنه های فیشینگ پاکسازی شده و از بین رفتند.فلذا مطالعه ما با مجوز از تارگت مورد حمله بوده است.

مراجع:

[1] OpenAI, “Introducing ChatGPT,” https://openai.com/blog/chatgpt.

[2] OpenAI, “GPT-3.5,” https://platform.openai.com/docs/models/gpt-3-5.

[3] P. Ray, “ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope,” *Internet of Things and Cyber-Physical Systems*, vol. 3, 04 2023.

[4] Nova, “How ChatGPT is Being Used in Virtual Assistants?” https://novaapp.ai/blog/chatgpt-virtual-assistant.

[5] HelpWise, “The complete Guide to Using Chat-GPT for Customer service,” https://helpwise.io/blog/how-to-use-chatgpt-for-customer-service.

[6] M. Daun and J. Brings, “How ChatGPT Will Change Software Engineering Education,” in *Proc. of the 2023 Conference on Innovation and Technology in Computer Science Education*, 2023, p. 110–116.

[7] J. McCoy, “ChatGPT for Content Creation: How to Write More in Less Time,” https://contentatscale.ai/chatgpt-for-content-creation/.

[8] Summa Linguae, “How to Use ChatGPT for Translation,” https://summalinguae.com/translation/how-to-use-chatgpt-for-translation.

[9] F. Huang, H. Kwak, and J. An, “Is ChatGPT Better than Human Annotators? Potential and Limitations of ChatGPT in Explaining Implicit Hate Speech,” in *Companion Proc. of the ACM Web Conference*, 2023.

[10] CheckPoint, "OPWNAI: Cybercriminals Starting to Use ChatGPT," https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/ .

[11] CyberARC, "Chatting Our Way Into Creating a Polymorphic Malware," https://www.cyberark.com/resources/threat-research-blog/ chatting-our-way-into-creating-a-polymorphic-malware, 2023.

[12] Meta, "The Malware Threat Landscape: NodeStealer, DuckTail, and More," https://engineering.fb.com/2023/05/03/security/malware-nodestealer-ducktail/, 2023.

[13] K. T. Gradon, "Electric Sheep on the Pastures of Disinformation and Targeted Phishing Campaigns: The Security Implications of ChatGPT," *IEEE Security & Privacy*, vol. 21, no. 3, pp. 58–61, 2023.

[14] Y. M. Pa Pa *et al.*, "An Attacker's Dream? Exploring the Capabilities of ChatGPT for Developing Malware," in *Proc. of CSET*. ACM, 2023.

[15] NameSheild, "ChatGPT, Can You Write a Phishing Email?" https://blog.nameshield.com/blog/2023/06/15/chatgpt-can-you-write-a-phishing-email/.

[16] S. S. Roy, K. V. Naragam, and S. Nilizadeh, "Generating Phishing Attacks using ChatGPT," 2023.

[17] A. Oest *et al.*, "Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale," in *Proc. of USENIX Security*, 2020, pp. 361–377.

[18] ——, "Inside a Phisher's Mind: Understanding the Anti-phishing Ecosystem Through Phishing Kit Analysis," in *Proc. of eCrime*, 2018.

[19] P. Zhang *et al.*, "CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing," in *Proc. of IEEE S&P*, 2021.

[20] S. Maroofi, M. Korczynski, and A. Duda, "Are You Human? Resilience ´ of Phishing Detection to Evasion Techniques Based on Human Verification," in *Proc. of ACM IMC*, 2020, pp. 78–86.

[21] B. Acharya and P. Vadrevu, "PhishPrint: Evading Phishing Detection Crawlers by Prior Profiling," in *Proc. of USENIX Security*, 2021.

[22] R. Liu *et al.*, "Inferring Phishing Intention via Webpage Appearance and Dynamics: A Deep Vision Based Approach," in *Proc. of USENIX Security*, 2022.

[23] S. Maroofi *et al.*, "COMAR: Classification of Compromised versus Maliciously Registered Domains," in *Proc. of IEEE Euro S&P*, 2020.

[24] A. Blum *et al.*, "Lexical Feature Based Phishing URL Detection Using Online Learning," in *Proc. of ACM AISec*, 2010.

[25] OpenAI, "Models," https://platform.openai.com/docs/models/gpt-4.

[26] ——, "OpenAI Codex," https://openai.com/blog/openai-codex .

[27] GitHub, “Your AI Pair Programmer,” https://openai.com/blog/openai-codex .

[28] A. J. ONeal, “Chat GPT “DAN” (and other “Jailbreaks”),”

https://gist.github.com/coolaj86/6f4f7b30129b0251f61fa7baaa881516.

[29] E. Fasllija *et al.*, “Phish-Hook: Detecting Phishing Certificates Using Certificate Transparency Logs,” in *SecureComm*, 2019.

منابع مکمل:

[1]. "How ChatGPT will change cybersecurity", [online] Available: https://www.kaspersky.com/blog/chatgpt-cybersecurity/46959/.

[2]. R. D. Caballar, "Cybercrime Meets ChatGPT: Look Out World", January 2023, [online] Available: https://spectrum.ieee.org/chatgpt-and-cybercrime.

[3]. A. Hughes, "ChatGPT: Everything you need to know about OpenAI's GPT-3 tool", January 2023, [online] Available: https://www.sciencefocus.com/future-technology/gpt-3/.

[4]. B. Gordijn and H. Have, "ChatGPT: evolution or revolution?", *Med Health Care and Philos*, 2023, [online] Available: https://doi.org/10.1007/s11019-023-10136-0.

[5]. F. Salahdine and N. Kaabouch, "Social Engineering Attacks: A Survey", *Future Internet*, vol. 11, no. 4, pp. 89, [online] Available: https://www.mdpi.com/1999-5903/11/4/89.

[6]. K. Chetioui, B. Bah, A. Ouali Alami and A. Bahnasse, "Overview of Social Engineering Attacks on Social Networks", *Procedia Computer Science*, vol. 198, pp. 656-661, 2022, [online] Available: https://doi.org/10.1016/j.procs.2021.12.302, ISSN 1877-0509.

[7]. L. Irwin, "The 5 Biggest Phishing Scams of All Time", October 2022, [online] Available: https://www.itgovernance.eu/blog/en/the-5-biggest-phishing-scams-of-all-time.

[8]. R. Zulfikar, "Phishing attacks and countermeasures" in Handbook of Information and Communication Security, Springer, 2010, ISBN 9783642041174.

[9]. A. Kumar Jain and B.B. Gupta, "A survey of phishing attack techniques defence mechanisms and open research challenges", *Journal of Enterprise Information Systems*, vol. 16, pp. 527-565, 2022, [online] Available: https://doi.org/10.1080/17517575.2021.1896786.

[10]. B. Gupta, N. Arachchilage and K. Psannis, "Defending against phishing attacks: taxonomy of methods current issues and future directions", *Telecommun Syst*, vol. 67, pp. 247-267, 2018, [online] Available: https://doi.org/10.1007/s11235-017-0334-z.

[**11**] J. Gori Mohamed, M. Mohammed Mohideen, N. Shahira Banu. E-Mail phishing an open threat to everyone. Int. J. Sci. Res. Publ. Volume 4, Issue 2, February 2014.

[**12**] Postfix. http://www.postfix.org.

[**13**] Air force mypers. http://www.afpc.af.mil/Support/myPers/.

[**14**] Google Home Added 600,000 More U.S. Users in 2018 Than Amazon Echo, But Amazon Echo Dot is Still the Most Owned Smart Speaker. https://voicebot.ai/2019/ 03/07/google-home-added-600000-more-u-s-users-in-2018-thanamazon- echo-but-amazon-echo-dot-is-still-the-most-owned-smart-speaker/.

[**15**] Alexa top 1 million websites. https://www.alexa.com/topsites.

[**16**] Alexa top sites by category. https://www.alexa.com/topsites/category.

[**17**] Alexa Skill Website. https://goo.gl/PvwL3T, Amazon Mechanical Turk. https:// www.mturk.com/.

[**18**] Amazon Alexa Skill Count Surpasses 30,000 in the U.S. https://voicebot.ai/ 2018/03/22/amazon-alexa-skill-count-surpasses-30000-u-s/.

[**19**] Amazon Alexa Skill Count Surpasses 80,000 worldwide. https://voicebot.ai/ 2018/03/22/amazon-alexa-skill-count-surpasses-30000-u-s/.

[**20**] Steps to test Amazon Alexa skill. https://developer.amazon.com/docs/ devconsole/test-your-skill.html#ways-to-test-an-alexa-skill.

[**21**] C. Jones, "The Top 10 Phishing Protection Solutions", January 2023, [online] Available: https://expertinsights.com/insights/top-10-phishing-protection-solutions/.

[**22**]. V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, et al., "A survey on smart grid potential applications and communication requirements", *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 28-42, Feb. 2013.

[**23**]**.** S. Garg, K. Kaur, G. Kaddoum, J. J. P. C. Rodrigues and M. Guizani, "Secure and lightweight authentication scheme for smart metering infrastructure in smart grid", *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3548-3557, May 2020.

[**24**]**.** C. Ang, *The Most Cyber Attsignificant Acks From 2006-2020 by Country*, 2021, [online] Available: https://www.visualcapitalist.com/cyber-attacks-worldwide-2006-2020/.

[**25**]**.** M. Benmalek and Y. Challal, "MK-AMI: Efficient multi-group key management scheme for secure communications in AMI systems", *Proc. IEEE Wireless Commun. Netw. Conf.*, pp. 1-6, Apr. 2016.

[**26**]**.** M. Z. Gunduz and R. Das, "Cyber-security on smart grid: Threats and potential solutions", *Comput. Netw.*, vol. 169, Mar. 2020.

[27]. R. E. Pérez-Guzmán, Y. Salgueiro-Sicilia and M. Rivera, "Communication systems and security issues in smart microgrids", *Proc. IEEE Southern Power Electron. Conf. (SPEC)*, pp. 1-6, Dec. 2017.

[28]. M. Z. Gunduz and R. Das, "Analysis of cyber-attacks on smart grid applications", *Proc. Int. Conf. Artif. Intell. Data Process. (IDAP)*, pp. 1-5, Sep. 2018.

[29]. H. Ritchie and M. Roser, *Two-Thirds of Global Population Will Live in Cities By 2050 Our World in Data*, 2018, [online] Available: https://ourworldindata.org/urbanization.

[30]. M. Benmalek, Y. Challal and A. Derhab, "Authentication for smart grid AMI systems: Threat models solutions and challenges", *Proc. IEEE 28th Int. Conf. Enabling Technologies: Infrastructure Collaborative Enterprises (WETICE)*, pp. 208-213, Jun. 2019.

[31]. A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi and R. Ahmad, "Machine learning and deep learning approaches for cybersecuriy: A review", *IEEE Access*, vol. 10, pp. 19572-19585, 2022.

[32]. A. Hasankhani, S. M. Hakimi, M. Bisheh-Niasar, M. Shafie-khah and H. Asadolahi, "Blockchain technology in the future smart grids: A comprehensive review and frameworks", *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021.

[33]. M. Ghiasi, T. Niknam, Z. Wang, M. Mehrandezh, M. Dehghani and N. Ghadimi, "A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past present and future", *Electr. Power Syst. Res.*, vol. 215, Feb. 2023.

[34]. M. Abdalzaher, M. Fouda, A. Emran, Z. Fadlullah and M. Ibrahem, "A survey on key management and authentication approaches in smart metering systems", *Energies*, vol. 16, no. 5, pp. 2355, Mar. 2023.

[35]. M. Sewak, S. K. Sahay and H. Rathore, "Deep reinforcement learning in the advanced cybersecurity threat detection and protection", *Inf. Syst. Frontiers*, vol. 25, pp. 589-611, Aug. 2022.

[36]. S. Banik, S. K. Saha, T. Banik and S. M. M. Hossain, "Anomaly detection techniques in smart grid systems: A review", *Proc. IEEE World AI IoT Congr. (AIIoT)*, pp. 331-337, Jun. 2023.

[37]. J. Kua, M. B. Hossain, I. Natgunanathan and Y. Xiang, "Privacy preservation in smart meters: Current status challenges and future directions", *Sensors*, vol. 23, no. 7, pp. 3697, Apr. 2023.

[38]. K. Y. Yap, H. H. Chin and J. J. Klemeš, "Blockchain technology for distributed generation: A review of current development challenges and future prospect", *Renew. Sustain. Energy Rev.*, vol. 175, Apr. 2023.

[39]. S. S. Koduru, V. S. P. Machina and S. Madichetty, "Cyber attacks in cyber-physical microgrid systems: A comprehensive review", *Energies*, vol. 16, no. 12, pp. 4573, Jun. 2023.

[40]. M. Lydia, G. E. P. Kumar and A. I. Selvakumar, "Securing the cyber-physical system: A

[41]. S. Bhattacharjee and S. K. Das, "Detection and forensics against stealthy data falsification in smart metering infrastructure", *IEEE Trans. Depend. Secure Comput.*, vol. 18, no. 1, pp. 356-371, Jan. 2021.

[42]. X. Lou, "Learning-based time delay attack characterization for cyber-physical systems", *Proc. IEEE Int. Conf. Commun. Control Comput. Technol. Smart Grids (SmartGridComm)*, pp. 1-6, Oct. 2019.

[43]. A. Ameli, A. Ayad, E. F. El-Saadany, M. M. A. Salama and A. Youssef, "A learning-based framework for detecting cyber-attacks against line current differential relays", *IEEE Trans. Power Del.*, vol. 36, no. 4, pp. 2274-2286, Aug. 2021.

[44]. S. Yankson and M. Ghamkhari, "Transactive energy to thwart load altering attacks on power distribution systems", *Future Internet*, vol. 12, no. 1, pp. 4, Dec. 2019.

The other side of the coin: the dark side of artificial intelligence; Practical cyber attack with artificial intelligence

Sina Manafi[1], Fardin Esmaili Sangri [2], Mehrdad Akbari [3]

Security Operations Center Researcher,

Sinahgfv@gmail.com

Computer Department, Urmia Branch, Islamic Azad University, Urmia, Iran,
Fardin.e.s62@gmail.com

Bachelor of Information Technology Engineering - Information Technology - Comprehensive University of Applied Sciences - Khaneh Kargar University Education Center, Tabriz, Tabriz, Iran, Email

Abstract— This paper explores the possibility of using ChatGPT to develop advanced phishing attacks and automate large-scale deployments. We make ChatGPT implement the following parts of a phishing attack: 1) impersonate a target website, 2) integrate code to steal user information, 3) obfuscate code, 4) automatically deploy the website to your hosting provider, 5) register phishing domain, and 6) integrating the website with a reverse proxy. It will make the initial evaluation of automatically generated phishing kits practical. It highlights their rapid production and deployment process as well as the close similarity of the resulting pages to the intended website. More broadly, we show that recent advances in artificial intelligence confirm the potential dangers of its misuse in phishing attacks, which could lead to an increase in their prevalence and severity. This highlights the need for advanced countermeasures and defensive measures in artificial intelligence systems.

Keywords: cyber attacks, artificial intelligence, ChatGPT, automation of attacks