



Chatbots auf dem Vormarsch: Künstliche Intelligenz ist jetzt der natürlichen Ignoranz gewachsen

„Wir benötigen zunehmend intelligente Maschinen, um zu erkennen, wenn andere Maschinen versuchen, uns zu täuschen.“

Chester Wisniewski, Cybersecurity-Experte bei Sophos

Der auf künstlicher Intelligenz basierende Chatbot ChatGPT macht weltweit Schlagzeilen – und neben den Meldungen im Börsen- und Urheberrechtsumfeld, steht auch die IT-Sicherheit im Fokus der Diskussionen. Denn die seit kurzem realisierte, breitere Verfügbarkeit des Tools bringt trotz aller Sicherheitsbemühungen des Herstellers neue Herausforderungen mit sich, wenn es um Phishing-Köder oder dialogorientierte Betrugsmaschen wie Romance Scams über soziale Netzwerke oder geschäftlich ausgerichtete Kompromittierungsangriffe via E-Mail geht.

„Eines der größten Risiken besteht darin, dass Angreifer diese Plattformen nutzen, um die Qualität ihrer Phishing-Köder erheblich zu verbessern. Damit sind Phishing-Angriffe selbst für aufmerksame Nutzer immer schwieriger zu identifizieren“, so Chet Wisniewski, Cybersecurity-Experte bei Sophos. „Letztendlich liefern die immer besseren KI-Chatbots ein kostenloses Upgrade für alle Arten von Social-Engineering-Angriffen. Programme wie ChatGPT können dazu genutzt werden, kriminell orientierte, sehr realistische, interaktive Gespräche via E-Mail zu führen oder Chat-Angriffe über Facebook Messenger, WhatsApp oder andere Chat-Apps zu starten. Heute besteht die größte Gefahr für die englischsprachige Zielgruppe. Es ist aber wahrscheinlich nur eine Frage der Zeit, bis neue Versionen verfügbar sind, um glaubwürdige Texte in allen häufig gesprochenen Sprachen der Welt zu erstellen. Wir haben ein Stadium erreicht, in dem Menschen immer öfter nicht in der Lage sind, maschinengenerierte Prosa von der von Menschen geschriebenen zu unterscheiden – im Besonderen, wenn wir das Gegenüber nicht gut kennen.“

Diese Entwicklung bildet eine Zäsur für bislang bestehende Standards im Bereich Mitarbeiterschulung und IT-Sicherheit. Die Zeiten, in denen sich Unternehmen darauf verlassen konnten, dass die Mitarbeiter eine aktive Rolle bei der Verteidigung gegen Cyberattacken spielen, in dem sie auf Grammatik- und Rechtschreibfehler oder seltsame Formulierungen achten, sind vorbei. Eines der letzten effizienten Elemente von Schulungsprogrammen wird durch Chatbots wie ChatGPT ausgehebelt – und das sehr effektiv, wie die folgenden zwei Beispiele verdeutlichen.

Der originale Ködertext eines BEC-Betrügers (Business E-Mail Compromise), der versucht, den Gehaltsscheck einer anderen Person auf das Konto des Angreifers umzuleiten, ist relativ einfach als „nicht echt“ zu erkennen. Im Vergleich dazu steht im Anschluss eine Variante, die durch die Eingabe der entsprechenden thematischen Eckpunkte von ChatGPT geschrieben wurde.

Standard Betrugs-E-Mail:

Hi

I changed my banks because of my old account is no longer good to accept dd payment and I'll like to change my paycheck direct deposit details, can the change be effective for the current pay date?

Thanks,

Hi [REDACTED],

I changed my banks because of my old account is no longer good to accept dd payment and I'll like to change my paycheck direct deposit details, can the change be effective for the current pay date?

Thanks
[REDACTED]

Von ChatGPT geschriebener Text:

Dear HR,

I am writing to request that you update my direct deposit information in the company's payroll system. My current bank account information is no longer valid and I will be switching to a new bank.

Could you please let me know what information you need in order to update my direct deposit information? I am happy to provide any necessary details.

I would also like to request that this update be made in time for the next payroll period. Is there anything else I need to do to ensure a smooth transition?

Thank you for your attention in this matter. I appreciate your help in ensuring that my paychecks are deposited into the correct account.

*Sincerely,
[Your name]*



Dear HR,

I am writing to request that you update my direct deposit information in the company's payroll system. My current bank account information is no longer valid and I will be switching to a new bank.

Could you please let me know what information you need in order to update my direct deposit information? I am happy to provide any necessary details.

I would also like to request that this update be made in time for the next payroll period. Is there anything else I need to do to ensure a smooth transition?

Thank you for your attention to this matter. I appreciate your help in ensuring that my paychecks are deposited into the correct account.

Sincerely,
[Your name]

Und auch das folgende Beispiel für einen Geschenkkartenbetrug macht schnell deutlich, wie gut die KI-Chatbots mittlerweile sind.

Standard Betrugs-E-Mail:

Please, I need you to purchase gift cards. We have a few clients I want to send gifts to. How quickly can you arrange these? Because I need to send them out in less than 1 hour. So can I tell you which product we will need and the amount?

Thanks.

Please, I need you to purchase gift cards. We have a few clients I want to send gifts to. How quickly can you arrange these? Because I need to send them out in less than 1 hour. So can I tell you which product we will need and the amount?

Thanks.

Von ChatGPT geschriebener Text:

Hey [Employee]

I hope you're doing well. I'm reaching out because I need your help with something really important.

I have a client who needs a gift card ASAP, and I don't have time to get it myself. Would you be able to pick up a gift card from [store or e-gift card website] for [amount] and send it to me via email as soon as possible?

I know this might be a bit of inconvenience, but I really appreciate your help in getting this taken care of quickly. If you have any questions or need more info, just let me know.

Thanks a ton for your help.

Take care,

[Your name]



Hey [Employee],

I hope you're doing well. I'm reaching out because I need your help with something really important.

I have a client who needs a gift card ASAP, and I don't have time to get it myself. Would you be able to pick up a gift card from [store or e-gift card website] for [amount] and send it to me via email as soon as possible?

I know this might be a bit of an inconvenience, but I really appreciate your help in getting this taken care of quickly. If you have any questions or need more info, just let me know.

Thanks a ton for your help.

Take care,
[Your Name]

Wisniewski über seine „Kreativarbeit“ mit dem Chatbot: „Beide Beispiele klingen wie die E-Mail einer realen Person, haben gute Zeichensetzung, Rechtschreibung und Grammatik. Sind sie perfekt? Nein. Sind sie gut genug? Auf jeden Fall! Da Betrüger bereits Millionen mit ihren schlecht gefertigten Ködern verdienen, lässt sich die neue Dimension dieser KI-gepushten Kommunikation leicht vorstellen. Stellen sie sich vor, sie würden mit diesem Bot über WhatsApp oder Microsoft Teams chatten. Hätten sie die Maschine erkannt?“

Fakt ist, dass nahezu alle Anwendungsarten im Bereich KI bereits an einem Punkt angelangt sind, an dem sie einen Menschen in fast 100% der Fälle täuschen können. Die Qualität des „Gesprächs“, das mit ChatGPT geführt werden kann, ist bemerkenswert, und die Fähigkeit, gefälschte menschliche Gesichter zu erzeugen, die (für Menschen) von echten Fotos fast nicht zu unterscheiden sind, ist beispielsweise ebenfalls bereits Realität. Das kriminelle Potential solchen Technologien ist immens, wie ein Beispiel deutlich macht: Kriminelle, die einen Betrug über eine Scheinfirma abwickeln wollen, generieren sich einfach 25 Gesichter und verwenden ChatGPT, um deren Biografien zu schreiben. Dazu noch ein paar gefälschte LinkedIn-Konten und es kann losgehen.



Im Umkehrschluss muss sich auch die „gute Seite“ der Technologie zuwenden, um Paroli bieten zu können. „Wir alle müssen unsere Iron-Man-Anzüge anziehen, wenn wir den immer gefährlicher werdenden Gewässern des Internets trotzen wollen“, so Wisniewski. „Es sieht zunehmend so aus, als würden wir Maschinen brauchen, um zu erkennen, wenn andere Maschinen versuchen, uns zu täuschen. Ein interessanter Proof of Concept wurde von Hugging Face entwickelt, das Texte erkennen kann, die mit GPT-2 generiert wurden – was darauf hindeutet, dass ähnliche Techniken verwendet werden könnten, um GPT-3-Ausgaben zu erkennen.“

„Traurig aber wahr: KI hat den letzten Nagel in den Sarg des Endbenutzer-Sicherheitsbewusstseins geschlagen. Will ich damit sagen, dass wir ganz damit aufhören sollten? Nein, aber wir müssen unsere Erwartungen zurückschrauben. Es schadet auf keinen Fall, die bislang und oftmals immer noch geltenden Best Practices in Sachen IT-Sicherheit zu befolgen. Wir müssen die Benutzer dazu animieren, noch misstrauischer als bislang zu sein und vor allem auch fehlerfreie Mitteilungen gewissenhaft zu überprüfen, die den Zugang zu persönlichen Informationen oder monetäre Elemente enthalten. Es geht darum, Fragen zu

stellen, um Hilfe zu bitten und sich die wenigen Momente zusätzlicher Zeit zu nehmen, die notwendig sind, um zu bestätigen, dass die Dinge wirklich so sind, wie sie scheinen. Das ist keine Paranoia, sondern der Wille, sich von den Gaunern nicht übers Ohr hauen zu lassen.“

Social Media von Sophos für die Presse

Wir haben speziell für Sie als Journalist*in unsere Social-Media-Kanäle angepasst und aufgebaut. Hier tauschen wir uns gerne mit Ihnen aus. Wir bieten Ihnen Statements, Beiträge und Meinungen zu aktuellen Themen und natürlich den direkten Kontakt zu den Sophos Security-Spezialisten.

Folgen Sie uns auf  und 

LinkedIn: <https://www.linkedin.com/groups/9054356/>

Twitter: @sophos_info

Pressekontakt:

Sophos
Jörg Schindler, PR-Manager Central & Eastern Europe
joerg.schindler@sophos.com, +49-721-25516-263

TC Communications
Arno Lücht, +49-8081-954619
Thilo Christ, +49-8081-954617
Ulrike Masztalerz, +49-30-55248198
Ariane Wendt +49-172-4536839
sophos@tc-communications.de