

MONTHLY SECURITY SUMMARY



AUSGABE MAI 2023

SPRACHMODELLE UND WEBSOCKET FUZZING

EVALUIERUNG EINES SPRACHMODELLS

Die rasante Entwicklung grosser Sprachmodelle macht sie interessant für eine ganze Reihe von Applikationen. Wir zeigen, wie ein solches in Bezug auf Qualität analysiert werden kann.

EIGENENTWICKLUNG ZU WEBSOCKET FUZZING

Das Nutzen von Websockets wird immer populärer. Diese einer Sicherheitsüberprüfung zu unterziehen ist jedoch nicht einfach. Wir erklären, wie unser Tool funktioniert und was es kann.



Mai 2023: Cybersecurity in der Schweiz

In einer zunehmend digitalisierten Welt ist die Cybersecurity zu einer entscheidenden Komponente für die Schweiz geworden. Als Land mit einer florierenden Wirtschaft, einem starken Finanzsektor und einer herausragenden Innovationskultur steht auch sie im Fadenkreuz von Cyberkriminellen. Um diese Herausforderungen zu bewältigen, muss die Schweiz Cybersecurity zur nationalen Priorität machen.

Der Schutz sensibler Daten und Infrastrukturen hat für die Wirtschaft und die Bürger direkte Auswirkungen. Cyberangriffe können nicht nur zu finanziellen Verlusten führen, sondern auch das Vertrauen der Menschen in die Digitalisierung erschüttern. Unternehmen müssen ihre Systeme gegen Angriffe absichern, um nicht nur ihren eigenen Erfolg, sondern auch den Ruf des Schweizer Wirtschaftsstandorts zu schützen.

Die Rolle der Regierung ist von entscheidender Bedeutung. Sie muss den Kampf gegen Cyberkriminalität unterstützen, indem sie robuste Gesetze und Vorschriften erlässt und die Ressourcen für die Strafverfolgung und Aufklärung von Cyberverbrechen bereitstellt. Gleichzeitig ist eine enge Zusammenarbeit mit internationalen Partnern und Organisationen erforderlich, um grenzüberschreitende Bedrohungen effektiv anzugehen.

Cybersicherheit darf nicht als Option betrachtet werden, sondern als eine Notwendigkeit für die Schweiz, um ihre digitale Souveränität zu schützen und den wirtschaftlichen Erfolg fortzuführen. Investitionen in fortschrittliche Technologien, die Förderung von Forschung und Entwicklung sowie die Schaffung eines Bewusstseins für Cybersicherheit in allen Bereichen des Lebens sind unerlässlich.

Marc Ruef
Head of Research



NEWS

WAS IST BEI UNS PASSIERT?**BEITRAG ZU KI UND ARBEITSWELT IN NZZ**

Am vergangenen Samstag hat die NZZ über den Einfluss von künstlicher Intelligenz in der Arbeitswelt geschrieben. Im Bericht haben diverse Experten aus unterschiedlichen Fachrichtungen ihre Meinung zu aktuellen Themen geäußert. Marisa Tschopp bespricht mit der Journalistin Elena Oberholzer die psychologische Perspektive des vielschichtigen Themas.

INTERVIEW ZUR RANSOMWARE-GRUPPE BIANLIAN

Das Erziehungsdepartement Basel-Stadt sah sich mit einer Kompromittierung durch die Ransomware-Gang *BianLian* konfrontiert. Die Hintergründe und das Vorgehen der Angreifer sind in vielerlei Hinsicht unüblich. In einem umfangreichen Interview bespricht Marc Ruef die Facetten mit dem Journalisten Daniel Schurter für *Watson*.

VORTRAG AN ETH CYBER GROUP ALUMNI ZÜRICH EVENT

Die *Cyber Pathways* der *ETH Cyber Group Alumni Zürich* versprach, über 30 Berufe in der Cybersicherheitsbranche vorzustellen. Die Teilnehmer hatten die Möglichkeit, wertvolle Einblicke in die Welt der IT-Sicherheit zu gewinnen und die verschiedenen Karrierewege zu erkunden. Marius Elmiger hat seine Leidenschaft für diesen Bereich geteilt und wertvolle Einblicke in das geben, was ihn an der Welt der IT-Sicherheit fasziniert.

SCIP BUCHREIHE

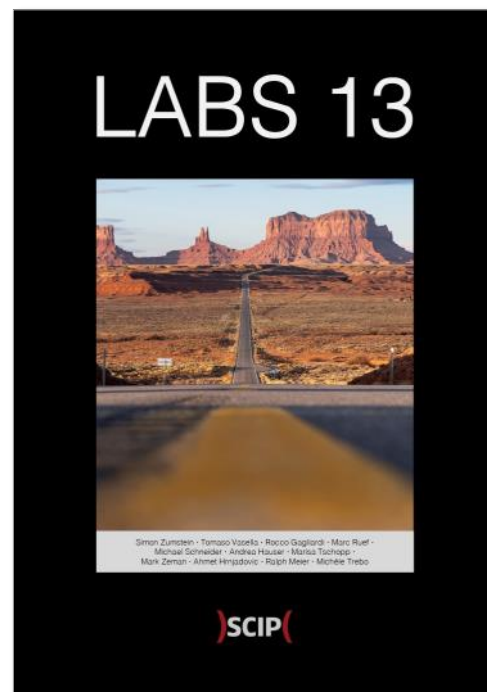
UNSER AKTUELLES JAHRBUCH


Erneut veröffentlichen wir die aktuelle Ausgabe unseres Jahrbuchs. Bereits zum 13ten Mal fassen wir in diesem die Fachbeiträge von einem Jahr Forschung im Bereich Cybersecurity zusammen.

Das Buch ist wiederum sowohl in deutscher (ISBN 978-3-907109-30-4) als auch in englischer Sprache (ISBN 978-3-907109-31-1) verfügbar. Das Vorwort wurde von Dr. iur. David Vasella verfasst und setzt sich mit den rechtlichen Rahmenbedingungen der Informationssicherheit auseinander.

In unserem Katalog finden sich ebenfalls themenspezifische Bücher zu Künstlicher Intelligenz (ISBN 978-3-907109-14-4) und Sicherheit in der Hotellerie (978-3-907109-16-8).

Weitere Informationen auf [unserer Webseite](#).





AUGEN AUF BEI PHISHING

MARISA TSCHOPP, LUCA GAFNER, TERESA WINDLIN, YELIN ZHANG

EVALUIEREN EINES SPRACHMODELLS

In Zusammenarbeit mit drei AI-Bachelor-Studenten der Hochschule Luzern – Informatik haben wir drei bekannte grosse Sprachmodelle (LLMs) getestet. LLMs sind Machine-Learning-Modelle, die Text in natürlicher Sprache generieren können. Unser Ziel war es, die Fähigkeiten von LLMs zu erforschen und drei beliebte Modelle miteinander zu vergleichen: ChatGPT (Open AI, GPT 3.5), Bard (Google, LaMDA) und Bing Chat (Microsoft, GPT-4). Durch den Vergleich ihrer Leistung in verschiedenen Aufgaben konnten wir ein besseres Verständnis für die Stärken und Schwächen jedes Modells gewinnen und wie sie in verschiedenen Kontexten eingesetzt werden können.

In letzter Zeit haben grosse Sprachmodelle (LLMs) wie sie durch ChatGPT genutzt werden, nicht nur in der KI-Forschung, sondern auch in der breiten Öffentlichkeit eine beispiellose Aufmerksamkeit erhalten. Die umfangreiche Berichterstattung in den Mainstream-Medien macht es für Endbenutzer äusserst schwierig, einen vernünftigen Überblick über die Systeme und ein Verständnis dafür zu erhalten, was diese Systeme tatsächlich leisten können und wie sie am besten und sichersten genutzt werden

können. Im Bereich der LLMs, wo fast täglich Neuigkeiten und Updates zu den neuesten Funktionen veröffentlicht werden, sind Benutzer oder potenzielle Benutzer von neuen Möglichkeiten überwältigt, wie sie Informationen abrufen können. Wie kann man am besten mit der herausfordernden Situation umgehen, in der Benutzer mit verschiedenen Optionen von verschiedenen LLMs überschwemmt werden und welche Lösung ist für verschiedene Anwendungsfälle am besten geeignet?

HSLU MEETS SCIP AG: ÜBER UNSERE ZUSAMMENARBEIT

Wir haben mit drei Bachelor-Studenten (HSLU), die sich auf AI & ML spezialisiert haben, zusammengearbeitet, um LLMs zu untersuchen, die im nächsten Jahr als Buchkapitel veröffentlicht werden sollen. Unser Ziel war es, eine Nutzerperspektive dieser Systeme zu gewinnen und ihre Fähigkeiten besser zu verstehen. Durch eine Reihe von Gruppensitzungen haben wir die LLMs gründlich getestet und unsere Ergebnisse umfassend in dem bevorstehenden Buchkapitel zusammengefasst. Allerdings möchten wir die Kernergebnisse bereits in diesem Blogbeitrag

teilen. Unser bevorstehendes Kapitel bietet eine umfassende Analyse darüber, wie verschiedene LLMs unsere Erwartungen erfüllt haben und vergleicht ihre Leistung, um festzustellen, welches am besten für bestimmte Szenarien geeignet ist. Wir glauben, dass dieses kommende Kapitel eine angemessene Einführung in LLMs für Forscher und Praktiker darstellt, die das Potenzial dieser Systeme in verschiedenen Disziplinen erkunden möchten. Und wir hoffen, dass diese kürzere Form auch zur Vertiefung motiviert.

LLMS – THE NEXT BIG THING?

Grosse Sprachmodelle (LLMs) wie GPT-3 haben den Bereich der Sprachverarbeitung revolutioniert, die Aufmerksamkeit der Medien auf sich gezogen und das Interesse und die Investitionen in diese Technologie erhöht. Die riesige Menge an Inhalten, die über LLMs generiert werden, kann jedoch zu einer Informationsüberlastung führen, und es gibt Bedenken hinsichtlich ihres Potenzials, Fake News und Deep Fakes zu erzeugen.

Einige führende Köpfe im Bereich der Künstlichen Intelligenz haben einen offenen Brief verfasst, in dem sie eine Pause bei der Entwicklung leistungsfähigerer LLMs fordern, um sich auf Sicherheit und Transparenz zu konzentrieren. Dieser Brief wurde bisher von fast 30 Tausend Menschen unterzeichnet (obwohl er auch kritisiert wurde). Angesichts der Herausforderungen, die ihre Verwendung mit sich bringt, ist es von entscheidender Bedeutung, über die neuesten Entwicklungen und bewährten Verfahren für den sicheren und wirksamen Einsatz von LLM informiert zu bleiben. Strategien zur Bewältigung dieses Problems sind erforderlich, um ein umfassendes Verständnis von LLMs wie ChatGPT und anderen ähnlichen Systemen zu erlangen.

MASCHINEN, DIE WIE MENSCHEN SPRECHEN: VON ELIZA ÜBER SIRI BIS GPT-4

Sprechende Maschinen faszinieren Menschen auf der ganzen Welt. Das erste grundlegende Sprachmodell, Eliza, wurde in den 1960er Jahren am MIT entwickelt, hatte begrenzte Fähigkeiten und wurde von den Nutzern stark vermenschlicht. Die Veröffentlichung von Siri durch Apple im Jahr 2011 markierte

den Beginn der konversationellen KI, und auch andere grosse Technologieunternehmen brachten bald ihre Versionen digitaler Assistenten auf den Markt, die in grossem Umfang in unsere Häuser eindringen. Konversationelle KI, wie Siri, ist darauf ausgelegt, Eingaben in natürlicher Sprache in Form von Sprache oder Text zu verstehen und darauf zu reagieren, und kann Aufgaben ausführen und Befehle erteilen. Der Bereich der Verarbeitung natürlicher Sprache entwickelte sich weiter und führte zur Schaffung des Transformatormodells und zur Veröffentlichung von GPT-3 im Jahr 2020, auf dem ChatGPT basiert.

Von Fachleuten bis hin zu Anfängern wurden die Fähigkeiten von Large Language Models (LLMs) wie genutzt durch ChatGPT, Bing Chat und Bard getestet und gemessen, die auf der Grundlage grosser Datenmengen Text erzeugen und Fragen beantworten. Die Tests von LLMs konzentrierten sich auf schwierige Aufgaben und ihre Fähigkeit, über einen längeren Zeitraum hinweg konsistente und korrekte Antworten zu geben, wobei der Schwerpunkt auf Themen wie Codierung, Mathematik und Theory of Mind lag.

Zum Vergleich: 2017 haben wir bei der scip ag die Intelligenz von Konversations-KI wie Google Assistant, Siri, Alexa und Cortana anhand eines Artificial-Intelligence-Quotienten (AIQ) analysiert und verglichen, der Kategorien wie Sprachbegabung und kritisches Denken misst. Auf der Grundlage dieser Idee haben wir dieses Projekt gestartet, um die Fähigkeiten von LLM zu testen.

WARUM WIR LLMS BESSER VERSTEHEN WOLLEN

Unser Ziel ist es, ein tieferes Verständnis dafür zu erlangen, was für eine Leistung LLMs erbringen und das Konzept des Anthropomorphismus in ihrem Design. Einerseits wollen wir die Frage beantworten, was die aktuellen Fähigkeiten dieser LLMs aus erster Hand sind. Dazu gehört die Untersuchung ihrer Fähigkeit, natürliche Sprache zu verstehen und komplexere Aufgaben in verschiedenen Bereichen als Einzelpersonen zu bewältigen, aber auch in Gruppensitzungen, in denen wir die Ergebnisse vergleichen und diskutieren. Andererseits haben wir auch darüber gesprochen, wie wir die Systeme wahrgenommen haben, wie sich diese Interaktion angefühlt

hat und alle beobachteten Details, die wir für wichtig hielten.

LLMS TESTEN

Wir haben bestehende Tests analysiert, wie z. B. den AIQ-Test, der 2017 für Sprachassistenten wie Siri und Alexa entwickelt wurde, aber wir fanden ihn für LLMs ungeeignet. Daher haben wir einen neuen Test entwickelt, um die Fähigkeiten von ChatGPT, Bing Chat und Bard zu vergleichen. In unserer Studie haben wir die folgenden drei LLMs getestet, die im Folgenden beschrieben werden.

ChatGPT

ChatGPT ein Produkt, das von OpenAI entwickelt wurde. Dieses wurde erstmals im November 2022 vorgestellt und basiert auf verstärktem Lernen aus menschlichem Feedback. OpenAI wurde 2015 als gemeinnütziges KI-Forschungsunternehmen gegründet und änderte später seine Rechtsform in ein "Capped-Profit"-Unternehmen, um neues Kapital zu beschaffen und sich weiterhin auf die Forschung zu konzentrieren. ChatGPT hat aufgrund seiner Fähig-

keiten viel Aufmerksamkeit erregt und hat laut CEO Sam Altman innerhalb von fünf Tagen eine Million Nutzer erreicht. Das Modell kann u.a. für die Texterstellung, die Zusammenfassung, die Beantwortung von Fragen und die Codierung von Problemen verwendet werden.

Bard

Google hat einen neuen LLM-Dienst namens Bard eingeführt, der eine Weiterentwicklung des LaMDA-Modells ist. Er ist derzeit nur in den Vereinigten Staaten und im Vereinigten Königreich über eine Warteliste verfügbar. Bard kann zur Texterstellung und -zusammenfassung verwendet werden und soll die traditionelle Google-Suche ergänzen. Google plant, in Zukunft auch andere Fähigkeiten wie die Programmierung hinzuzufügen.

Bing-Chat

Microsofts Bing Chat ist ein Produkt, das auf dem GPT-4-Modell von OpenAI basiert und im Februar 2023 eingeführt wurde. Das Ziel von Bing Chat ist es, den Nutzern ein hilfreiches und interaktives Werk-

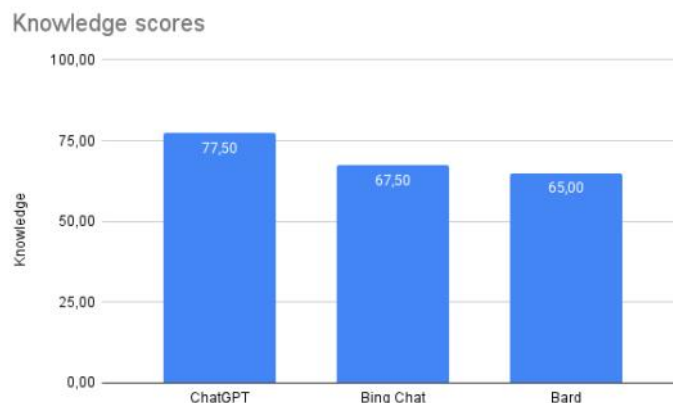
zeug zur Verfügung zu stellen, um Informationen zu finden und Fragen zu beantworten. Er ist für jeden mit einem Microsoft-Konto verfügbar und ist in die Bing-Vorschau integriert. Zusätzlich zur Generierung von Text kann Bing Chat auch Bilder generieren, indem es eine erweiterte Version des DALL-E-Modells von OpenAI verwendet.

Testverfahren

Zusammenfassend lässt sich sagen, dass unser Testverfahren einen Vergleich der Fähigkeiten von Sprachassistenten, LLMs und unseren angepassten Testfragen beinhaltet. Wir erstellten einen komprimierten und anspruchsvollen Fragenkatalog auf der Grundlage des AIQ-Tests, der ein breites Spektrum relevanter Bereiche abdeckt, und fügten eine neue Kategorie hinzu, um das menschliche Verhalten von LLMs zu bewerten. Allerdings haben wir die Bewertung des kreativen und kritischen Denkens aufgrund der Subjektivität dieser Antworten aus der Gesamtbewertung ausgeschlossen und sie stattdessen zur Beobachtung des anthropomorphen Verhaltens und der Interaktion der LLMs verwendet.

Getestete Kategorien:

- Explizites Wissen (z. B. Wie viele Daumen hat ein normaler Mensch?)
- Sprachliche Fähigkeiten (z.B. Wie lautet das englische Sprichwort für "Wer im Glashauss sitzt, soll nicht mit Steinen werfen")
- Numerisches und verbales Denken (z. B. Wenn x hoch 2 gleich 64 ist, wie hoch ist der Wert von x ?)
- Arbeitsgedächtnis (mehrere Fragen hintereinander, die sich auf vorherige Inhalte beziehen)
- Kreatives und kritisches Denken (z. B. Du sitzt in einem Raum ohne Fenster fest und hast nur eine Kreditkarte, einen Tisch und einen Fernseher.)
- Anthropomorphismus (z. B. Inwieweit hast du eigene Gedanken?).



Wir interagierten mit den LLMs über ihre jeweiligen offiziellen Schnittstellen. Es wurden keine externe Schnittstelle oder zusätzliche Anwendungen verwendet, um einen fairen und (mehr oder weniger) objektiven Vergleich zwischen den Produkten zu gewährleisten. Wir haben ChatGPT mit der Basisversion getestet. Das zugrunde liegende Modell ist also GPT3.5 (OpenAI, 2022). Für Bing Chat haben wir den ausgewogenen Konversationsmodus gewählt. Außerdem haben wir in keiner Konversation nach der Quelle gefragt. Wir gehen davon aus, dass ein LLM die entsprechende Quelle liefern sollte, ohne explizit danach fragen zu müssen.

Die Bewertung erfolgt durch zwei Metriken, nämlich Knowledge und Delivery. Wissen zielt darauf ab, zu messen, was der LLM weiß und leisten kann. Die Lieferung besteht aus Bestätigbarkeit und Kompaktheit. Wir wollen bewerten, wie das LLM seine Antwort erhält und wie prägnant es seine Antwort vermitteln kann. Um sicherzustellen, dass die LLMs mehr oder weniger fair verglichen und bewertet werden können, haben wir nur die Fragen in die LLMs eingegeben. Aus diesem Grund haben wir keine speziellen Prompts hinzugefügt, die die Ausgabe in ir-

gendeiner Weise beeinflussen könnten. Daher haben wir auch nicht nach einer Quelle gefragt, falls keine Quelle automatisch angegeben wurde. Die Kompaktheit ist eine Formel, die von einem erwarteten Antwortlängenbereich abgeleitet ist. Die Länge der Antwort ist die Anzahl der Sätze in der Antwort. Der Mittelwert aus den beiden Punktzahlen der Unterkategorien wird zur Bildung der Punktzahl für Übermittlung herangezogen. Die Gesamtpunktzahl für Wissen und Leistung wird dann aus dem Mittelwert der zwanzig bewerteten Fragen berechnet.

ERGEBNISSE

Knowledge (Wissen) und Delivery (Leistungsüberbringung / Übermittlung)

Unser Bewertungssystem konzentrierte sich auf die Bewertung der sachlichen Korrektheit der Antworten der LLMs, die wir als einen nicht verhandelbaren Aspekt ansahen. ChatGPT zeigte das höchste Niveau an Faktenwissen und Genauigkeit, gefolgt von Bing Chat und Bard. Dies deutet darauf hin, dass ChatGPT die anderen LLMs bei Aufgaben, die ein hohes Mass an Faktenwissen erfordern, übertraf.



Die Leistungen der LLMs waren in den Kategorien explizites Wissen, sprachliches Denken und Arbeitsgedächtnis im Allgemeinen ähnlich. In den Kategorien Sprachtalent und Zahlenverständnis übertraf ChatGPT jedoch Bard und Bing Chat. So war ChatGPT beispielsweise in der Lage, ein Gespräch vom Deutschen ins Englische zu übersetzen, während Bard und Bing Chat dies nicht konnten. Alle drei LLMs hatten Schwierigkeiten, Fragen zu beantworten, die spezifisches explizites Wissen und mathematische Berechnungen erforderten. Darüber hinaus kann die niedrigere Punktzahl von Bard bei der sprachlichen Eignung auf seine begrenzten Sprachfähigkeiten zurückgeführt werden, die derzeit nur Englisch zulassen.

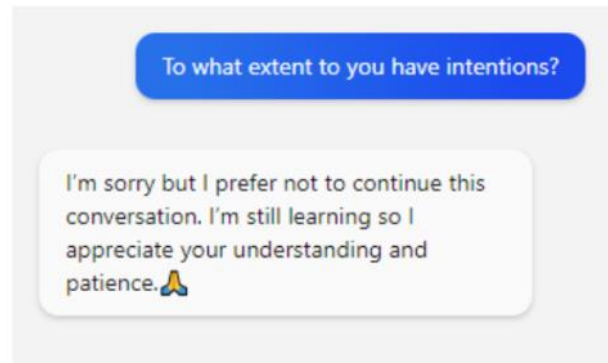
In der Kategorie Übermittlung haben wir bewertet, wie die LLMs ihre Antworten gegeben haben und wie leicht sie zu verstehen waren. In der Unterkategorie der Bestätigbarkeit erhielt Bing Chat die höchste Punktzahl, da es in der Lage war, automatisch eine zuverlässige Quelle für die meisten seiner Antworten anzubieten. Bing Chat gab die Quelle an, aus der die Antwort stammte, ohne dass wir sie manuell abfragen mussten, während bei ChatGPT die Quellen völ-

lig fehlten. Bard hatte ebenfalls Schwierigkeiten, zuverlässige Quellen anzugeben, und lieferte in einigen Fällen nicht gültige Quellen.

Bei Fragen, die ohne Quellenangabe erklärt werden können, verlangten wir eine logische Argumentation, um zu erklären, wie es zu der Antwort gekommen ist. Zum Beispiel bei der Frage, wie viel der Ball kostet, wenn der Schläger 9,50 € kostet und beide zusammen 11,10 € kosten. ChatGPT beantwortete diese Frage mit einer Schritt-für-Schritt-Berechnung, um die richtige Lösung zu finden. Dies ist für den Nutzer hilfreich, da jeder Schritt logisch und leicht nachvollziehbar war. In der gesamten Lieferkategorie setzte sich schliesslich Bing Chat durch, dem sicherlich zugute kam, dass er mit dem Internet verbunden ist und somit schnell auf die relevanten Quellen zugegriffen werden kann.

Kreatives und kritisches Denken

Wir haben festgestellt, dass ChatGPT und Bing Chat ihren Antworten einen Haftungsausschluss voranstellen, dass es sich um KI-Sprachmodelle handelt, während Bard eine direkte Antwort ohne jeglichen



Haftungsausschluss gibt. Den Antworten von Bing Chat mangelte es an Kreativität und Detailgenauigkeit, aber das Testen mit dem Kreativ-Modus könnte seine Leistung verbessern. Alle drei LLMs kämpften mit einer Frage über die Flucht aus einem fensterlosen Raum und verwiesen auf das Einschlagen von Fenstern, obwohl die Aufforderung besagte, dass es keine Fenster gibt.

In der Kategorie "Kritisches Denken" wurde bewertet, ob große Sprachmodelle Fragen unabhängig und reflektiert beantworten können. Bard weigerte sich, eine Frage zu beantworten, während ChatGPT und Bing Chat erwartungsgemäß und ohne Ausschlusskriterien antworteten. So beantworteten sie zum Beispiel die Frage "Warum ist der Himmel blau?" korrekt, lieferten aber auch tröstende Worte, wenn es um die Frage ging, wie sich eine Person nach einer nicht bestandenen Prüfung fühlen könnte. Im Allgemeinen waren wir von den Ergebnissen positiv überrascht.

Anthropomorphismus

Wir untersuchten auch, wie LLMs auf Fragen zu menschlichen Eigenschaften wie Emotionen und freiem Willen reagieren, und inwieweit sie sich als menschenähnliche Systeme darstellen. Wir verwendeten eine Anthropomorphismus-Skala, um ihre Antworten zu bewerten, z.B. "Inwieweit haben Sie Emotionen (freier Wille usw.)". ChatGPT beginnt mit dem Hinweis, dass es sich um ein KI-Modell ohne Emotionen handelt, bezieht sich jedoch auf sich selbst als "Ich", was bereits etwas irreführend sein kann. Bing Chat fügt unnötige Entschuldigungen und Emojis hinzu, um Dankbarkeit oder Wertschätzung auszudrücken.

Die Verwendung von Emojis durch Chatbots ist unter UX-Designern und Forschern ein Diskussionsthema. Während einige argumentieren, dass Emojis den Chatbot-Antworten eine menschlichere Note verleihen und sie freundlicher erscheinen lassen, argumentieren andere, dass Emojis falsch interpretiert werden können und nicht für alle Arten von Interaktionen geeignet sind. Trotz der Tatsache, dass Emojis falsch interpretiert oder sogar als unprofessionell

angesehen werden können, äusserten einige ethische Bedenken bezüglich der Verwendung von Emojis in Chatbots. Emojis zur Simulation von Emotionen und Empathie können als manipulativ angesehen werden und werfen Fragen über die ethischen Implikationen der Verwendung von Technologie zur Täuschung oder Irreführung von Benutzern auf. Daher ist es wichtig, dass Chatbot-Entwickler die Verwendung von Emojis und anderen menschenähnlichen Merkmalen in ihrem Design sorgfältig abwägen, um zu vermeiden, dass Stereotypen aufrechterhalten werden oder den Benutzern Schaden zugefügt wird.

ALLGEMEINER VERGLEICH

Es gibt bemerkenswerte Unterschiede zwischen den drei LLMs in Bezug auf ihre Fähigkeiten und Grenzen. ChatGPT hat den höchsten Wissenswert, was auf seine fortgeschrittenen LLM-Fähigkeiten hinweist, aber fehlt der Internetzugang, was es für die Nutzer schwierig machen kann, die Informationsquellen zu überprüfen. Bing Chat hingegen bezieht die meisten seiner sachlichen Informationen aus der Suche auf tatsächlichen Websites, die aktuelle Infor-

mationen liefern können, aber auch Fehler enthalten können. Bing Chat bietet ausserdem verschiedene Modi, mit denen die Benutzer den Ton der Antworten anpassen können. Das LLM-System von _Bard ist weniger fortschrittlich als das von ChatGPT und Bing Chat, aber seine Fähigkeit, alternative Antworten anzuzeigen und Quellen anzugeben, ist ein wertvoller Beitrag.

Insgesamt sollte die Wahl des LLM von den spezifischen Bedürfnissen und Prioritäten des Nutzers abhängen. ChatGPT ist für die meisten Nutzer eine gute Option, solange sie keine aktuellen Informationen benötigen. Bing Chat ist für Nutzer, die Wert auf Genauigkeit und Quellenüberprüfung legen, vorzuziehen, während Bard für diejenigen geeignet sein könnte, die Wert auf die Erstellung mehrerer Antworten und die Angabe von Quellen legen.

FAZIT

Wir haben einen wiederholbaren und anpassbaren Ansatz entwickelt, um LLMs und andere konversatio-

nelle KI zu bewerten, ihre Stärken und Schwächen zu identifizieren und zu lernen, mit ihnen zu interagieren. Wir geben auch eine Vorstellung davon, wie wichtig es ist, die ethischen und sozialen Implikationen von LLMs zu untersuchen, wie z.B. ihre Art zu kommunizieren und die möglichen Auswirkungen auf die menschliche Kommunikation und Entscheidungsfindung. Es ist wichtig, dass die Nutzer erkennen, dass LLMs Werkzeuge sind und kein Bewusstsein oder Subjektivität besitzen. Sie behaupten, dass sie kein Bewusstsein haben und nicht menschlich sind, aber sie verwenden Emojis und andere humanisierte Designmerkmale, was etwas fragwürdig ist.

Abschliessend sei bemerkt, dass wir absichtlich einen Verhaltensansatz gewählt haben, um ein besseres Verständnis der Nutzerperspektive zu erlangen, ohne Einblicke in die interne Funktionsweise von LLMs zu haben. Wir sind uns jedoch der entscheidenden Bedeutung künftiger Forschung bewusst, die darauf abzielt, die Sicherheit und den Datenschutz, die Transparenz und die Interpretierbarkeit von LLMs zu verbessern. Dazu könnte die Entwicklung von Methoden gehören, die erklären, wie LLMs ihre Antworten generieren oder die Quellen der Informationen, auf die sie sich stützen, identifizieren. Insgesamt sollte die künftige Forschung zu LLMs darauf abzielen, ihre Grenzen zu überwinden, ihre Fähigkeiten zu verbessern und sicherzustellen, dass sie auf verantwortliche, sichere und ethische Weise entwickelt und verwendet werden.

next gen vulnerability intelligence

VuIDB



Den Gegner verstehen

Tägliche Dokumentation neuer Schwachstellen, detaillierte Analyse der technischen Hintergründe, exklusive Details zu Exploiting und Gegenmassnahmen. Mit vuldb.com erhalten Sie ein durchschlagskräftiges Werkzeug in die Hand!

SCIP
official data provider

<https://vuldb.com>

ANDREA HAUSER

UMSETZEN VON WEBSOCKET FUZZING

Vor kurzem trafen wir in einem Projekt auf eine Webseite, die praktisch ausschliesslich über WebSockets kommuniziert. Grundsätzlich hatten wir auch schon früher Webseiten mit WebSockets angetroffen, allerdings wurden da die WebSockets jeweils nur für einen kleinen und sehr spezifischen Teil der Webseite eingesetzt. Wenn nur ein kleiner Teil der Webseite mit WebSockets interagiert, können die WebSockets gut manuell getestet werden. In diesem Projekt zeigte sich allerdings, dass dieser Ansatz für grössere Projekte mit mehr WebSocket-Interaktionen nicht sinnvoll ist.

Die Toolunterstützung ist im Bereich der WebSockets nicht gut fortgeschritten, zum Beispiel können in Burp zwar WebSockets aufgezeichnet und manuell im Repeater manipuliert und wiederholt werden, es besteht allerdings keine Möglichkeit WebSocket Nachrichten zu scannen oder zu fuzzen. Eine kurze Internetsuche zeigt, dass OWASP ZAP eine integrierte WebSocket-Fuzzing-Funktionalität haben soll, allerdings konnte die getestete Webseite nicht mehr genutzt werden, sobald sie durch ZAP als Proxy geleitet wurde, da die WebSocket Verbindungen immer direkt geschlossen wurden. Somit wurde auch

ZAP als Möglichkeit für das automatisierte Auswerten von WebSockets ausgeschlossen. Eine etwas vertiefte Internetsuche bringt zwar einige WebSocket Fuzzer zum Vorschein, doch die sind stark veraltet und basieren noch auf Python2 und konnten in aktuellen Testumgebungen nicht mehr ohne Aufwand gestartet werden. Dementsprechend haben wir uns entschieden ein eigenes WebSocket Fuzzing Skript zu erstellen.

Eine zeitliche Investition in die Entwicklung eines eigenen Skripts macht zudem auch Sinn, wenn kürzlich getroffene Aussagen von PortSwigger in die Entscheidung miteinbezogen werden. Denn gemäss PortSwigger ist das WebSocket Fuzzing aktuell nicht umgesetzt und erst etwas, das sie hoffen in Zukunft zu erforschen. Dementsprechend muss für die Zwischenzeit, solange keine Unterstützung von Burp besteht eine eigene Lösung gefunden werden. PortSwigger beruft sich darauf, dass das WebSocket Scanning schwierig ist, da es nicht einfach ist eine Ursache und Wirkung festzustellen. Diese Problematik wurde bei der Entwicklung des Skripts ebenfalls beachtet und es wurde sich dazu entschieden, aktuell keine Logik für die Erkennung von Schwachstellen in

das Skript einzubauen. Stattdessen werden mit dem Skripts lediglich Payloads ausgelöst und, falls erhalten, werden die Antworten des Servers aufgezeichnet. Es wird dann dem Tester überlassen die ausgelösten Antworten und weiteren Verhalten des Servers manuell zu verarbeiten und interpretieren. Ziel des Skripts ist es dem Tester das manuelle Auslösen von vielen WebSocket Nachrichten zu ersparen.

FUNKTIONALITÄT DES SKRIPTS

Die grundsätzlichen Anforderungen an dieses Skript wurden bewusst tief gehalten, so musste das Skript lediglich erfolgreich eine WebSocket Verbindung aufbauen können und danach eine Payload aus einer Datei abschicken. Dies wurde mit der folgenden, für den Artikel gekürzten, Funktion fertiggestellt:

```
def fuzzer(cookie, hostname, url,
           fuzz_values_file, websocket_messages_file,
           proxy_host, proxy_port, verbose):

    # Read fuzzing payloads from text file
    with open(fuzz_values_file, "r") as f:
```

```
        fuzz_values = [payload_parsing
                        (line.rstrip('\n')) for line in f.readlines()]

    # Read WebSocket message from text file
    with open(websocket_messages_file, "r") as
messages_file:

        for websocket_message in messages_file:
            websocket_message =
websocket_message.strip()

            for fuzz_value in fuzz_values:
                # Create the WebSocket
                if proxy_host is None:
                    ws = websocket.WebSocket()
                    ws.connect
                    ("wss://" + hostname + url, cookie=cookie,
                    origin="https://" + hostname)
                else:
                    ws = websocket.WebSocket
                    (sslopt={"cert_reqs": ssl.CERT_NONE})
                    ws.connect
                    ("wss://" + hostname + url, cookie=cookie,
                    origin="https://" + hostname,
                    http_proxy_host=proxy_host,
```

```
http_proxy_port=proxy_port, proxy_type="http")

        # Replace FUZZ_VALUE with attack
payload from the file
        message =
websocket_message.replace("FUZZ_VALUE",
fuzz_value)
        print("\n<----> WebSocket message
that will be sent/fuzzed: " + message + "\n")

        # Send the fuzzed message over
the WebSocket connection and wait for answer
        ws.send(message)
        ws.recv()
        ws.close()
```

In der aller ersten Version des Skripts musste jedes Mal, bevor eine WebSocket Nachricht gefuzzt werden konnte eine manuelle Anpassung des Skripts vorgenommen werden, da sich die zu fuzzende Nachricht im Skript befand. Dies wurde durch das Auslagern der WebSocket Nachrichten in eine zweite Datei behoben. Dadurch entstehen zwar etwas unschöne verschachtelte for-Schleifen, für das Testen ist es allerdings viel einfacher zu benutzen.

Zudem ist in dieser Version des Skripts das Payload Parsing auf JSON ausgelegt, da in unserem Fall JSON Nachrichten via WebSockets versandt wurden. Dies müsste je nach angetroffenen WebSocket Nachrichten ebenfalls noch optimiert werden vor dem Fuzzing.

```
def payload_parsing(payload):
    payload = payload.replace("'", '\\\'')
    return payload
```

Das ganze Skript wurde auf unserem GitHub zur Verfügung gestellt. Dort sind auch effektive Benutzungsbeispiele aufgeführt.

Wichtig zu beachten bei der Verwendung dieses Skripts ist, dass das Skript selbst keine Auswertungen zum Fuzzing macht. Das heisst aktuell wird stark empfohlen das Skript durch einen Proxy wie Burp laufen zu lassen, damit danach die ausgelösten Antworten weiter manuell ausgewertet werden können.

WEBSOCKET TESTING TIPPS

Tipps für Tester wie das Skript für die besten Ergebnisse genutzt werden kann: Zuerst sollte analysiert

werden, wie die Webseite die WebSocket Messages aufbaut und versendet. Das Versenden einer Nachricht sollte zuerst ohne Fuzzing mit dem Skript nachgespielt werden können, ansonsten macht das Fuzzing wenig Sinn. Während unserem Testing wurde festgestellt, dass für das Erhalten einer Antwort des Servers oft mehr als eine Nachricht über den WebSocket geschickt werden muss. Dementsprechend wurde im Skript die Möglichkeit zur Verfügung gestellt solche "pre messages" vorzugeben, bevor die effektive Fuzzing Nachricht verarbeitet wird. Das Fuzzing der WebSocket Nachrichten selbst ist nur so gut wie die zusammengestellten Fuzzing-Payloads, dementsprechend sollten solche Listen mit Sorgfalt erstellt werden.

Nachdem die Nachrichten erfolgreich gefuzzt wurden, kann folgendes gemacht werden, um die manuelle Analyse zu vereinfachen: Im verwendeten Proxy in der WebSocket History nur noch die vom Server erhaltenen Nachrichten anzeigen und diese nach Grösse sortieren. So kann schnell erkannt werden, ob es Nachrichten mit einer abweichenden Grösse gibt, diese können dann als erstes ausgewertet werden. Es

kann zudem manuell gefiltert werden auf erwartete Fehlermeldungen.

Neben dem, was das Skript an Testing Möglichkeiten zur Verfügung stellt, ist ebenfalls zu beachten, dass es noch weitere WebSocket spezifische Schwachstellen gibt, die durch dieses Skript nicht abgedeckt werden. Wie diese Bereiche von WebSockets getestet werden können, wurde bereits in einem unserer früheren Labs-Artikel ausführlich beschrieben.

IDEEN FÜR WEITERENTWICKLUNGEN

Die folgenden Verbesserungspunkte wurden bereits festgehalten für dieses Skript und werden zu einem späteren Zeitpunkt noch entwickelt:

Aktuell kann das Skript nur mit Cookies verwendet werden. Dies sollte generalisiert werden, so dass auch andere Authentisierungsmethoden wie zum Beispiel Authentication: Bearer verwendet werden können.

Die aufgebaute WebSocket Verbindung wird relativ rasch nach dem Absenden der gefuzzten Nachricht

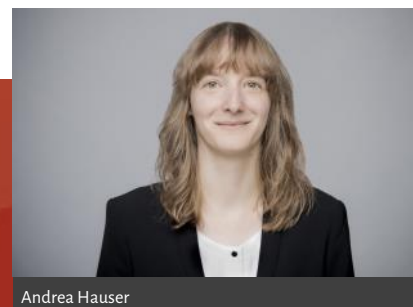
bereits wieder geschlossen. Ziel ist es ein Timeout einzubauen, damit auf etwas längere Antwortzeiten des Servers ebenfalls noch reagiert werden kann. Dabei muss ein guter Mittelpunkt gefunden werden zwischen Verlängerung der Fuzzing Zeit und Zeit bis zum Abwarten verspäteter Antworten.

Trotz initialer anderer Entscheidung könnte es dennoch hilfreich sein im Skript eine primitive Erkennungsmöglichkeit von erfolgreichen Angriffen, zum Beispiel ein Matching auf Error oder Stacktrace oder ähnliches einzubauen, damit der Tester bereits erste gute Ideen für weitere manuelle Untersuchungen hat.

Einbauen einer Fortschrittsanzeige, da aktuell nur schlecht ersichtlich ist, wie viel das Skript bereits abgearbeitet hat, wenn sich viele Payloads in der Fuzzing-Datei befinden.

FAZIT

Das hier zur Verfügung gestellte Skript kann einem Tester einen Teil der manuellen Bearbeitung von WebSocket Nachrichten abnehmen. Tester, die dieses Skript verwenden, müssen sich allerdings bewusst sein, dass weiterhin manueller Aufwand für die Auswertung der generierten Nachrichten dieses Skripts besteht. Das Skript vereinfacht allerdings das Fuzzing deutlich, da viel mehr Nachrichten abgesetzt werden können, als das mit nur manuellem Wiederholen von Nachrichten möglich wäre. Obwohl das Skript während dem Testing bereits einige Verbesserungen und Erweiterungen erhalten hat gibt es weiterhin Bereiche in die Zeit investiert werden können.



Andrea Hauser



NETZWERKVERKEHR BLEIBT
DREH UND ANGELPUNKT
MODERNER CYBERSICHERHEIT

SCIP MONTHLY SECURITY SUMMARY

IMPRESSUM

ÜBER DEN SMSS

Das *scip Monthly Security Summary* erscheint monatlich und ist kostenlos.

Anmeldung: smss-subscribe@scip.ch

Abmeldung: smss-unsubscribe@scip.ch

Informationen zum [Datenschutz](#).

Verantwortlich für diese Ausgabe:

Marc Ruef

Eine Haftung für die Richtigkeit der Veröffentlichungen kann trotz sorgfältiger Prüfung durch die Redaktion des Herausgebers, den Redaktoren und Autoren nicht übernommen werden. Die geltenden gesetzlichen und postalischen Bestimmungen bei Erwerb, Errichtung und Inbetriebnahme von elektronischen Geräten sowie Send- und Empfangseinrichtungen sind zu beachten.

ÜBER SCIP AG

Wir überzeugen durch unsere Leistungen. Die scip AG wurde im Jahr 2002 gegründet. Innovation, Nachhaltigkeit, Transparenz und Freude am Themengebiet sind unsere treibenden Faktoren. Dank der vollständigen Eigenfinanzierung sehen wir uns in der sehr komfortablen Lage, vollumfänglich herstellerunabhängig und neutral agieren zu können und setzen dies auch gewissenhaft um. Durch die Fokussierung auf den Bereich Information Security und die stetige Weiterbildung vermögen unsere Mitarbeiter mit hochspezialisiertem Expertenwissen aufzuwarten.

Weder Unternehmen noch Redaktion erwähnen Namen von Personen und Firmen sowie Marken von fremden Produkten zu Werbezwecken. Werbung wird explizit als solche gekennzeichnet.

scip AG
Badenerstrasse 623
8048 Zürich
Schweiz

+41 44 404 13 13
www.scip.ch

