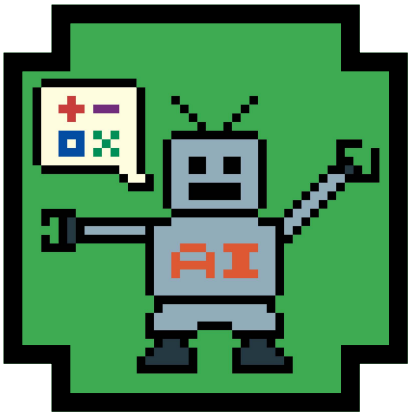
AI Chatbots: *Benefits and Risks*

Office of Information Technology
Division of Information Security
NOVEMBER 2024

As our work environment steadily moves towards using Artificial Intelligence (AI), chatbots are becoming increasingly popular in customer interactions, online queries, and problem-solving. AI chatbots, such as ChatGPT by Open AI, are fast becoming a basic tool in digital communication from customer service to personal assistants, but so are the risks associated with their use. Understanding the risks (which may sometimes outweigh chatbot benefits) can ensure cybersecurity and personal safety for all of us. The November IHS Cybersecurity Awareness Newsletter discusses some advantages and risks of using AI chatbots and alerts us of our new online reality.

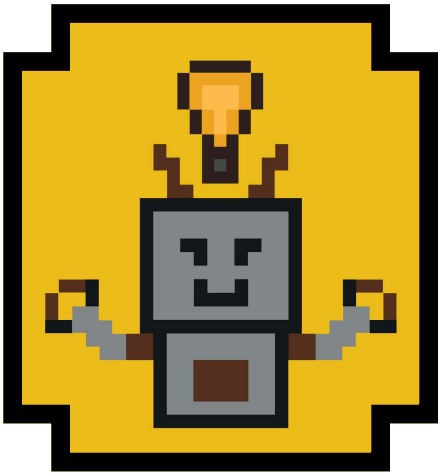


Benefits of AI Chatbots

So far, most of us have discovered some chatbot benefits, whether searching for products or a service or just seeking an answer to a question. Chatbots pop up to assist in meeting online searches. Most chatbot answers provide a wealth of information. Therefore, some obvious benefits of chatbot uses are that they reduce human error, especially in tasks involving repetitive data entry or processing, improve user engagement in natural, conversational interactions; efficiently collect data as chatbots are data-driven tools that can collect customer feedback and analyze insights to improve user experience; they offer multilingual support by communicating in multiple languages; they are scalable, that is, as a search grows chatbots can handle thousands of requests without the need for live agents; they offer personalized experiences with user data to provide tailored recommendations and suggestions. These are just some benefits of chatbots.

Understanding the Risks of Chatbots

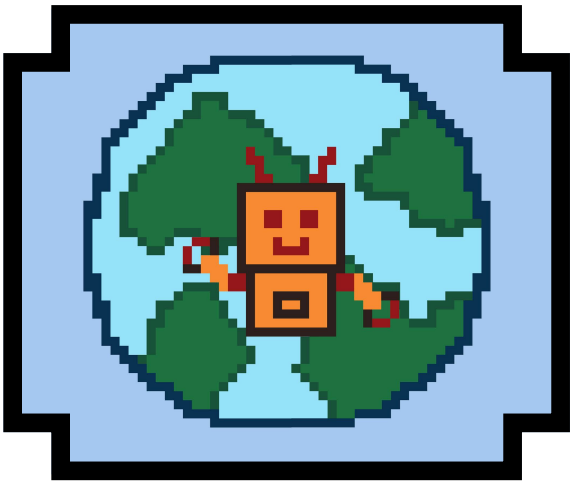
However, on the flip side, there are risks associated with chatbots, such as ChatGPT, which can be broadly categorized into three general areas: misinformation and disinformation, social engineering attacks, and privacy concerns.



Misinformation and Disinformation: Chatbot training relies on massive datasets of text and code, much of which could contain biased or inaccurate information. This can result in misleading or false responses, such as incorrect financial advice or rumors spread by conspiracy theorists.

Social Engineering Attacks: This type of attack takes advantage of the trust humans put into the online accounts they interact with, leaving a door open for attackers to leverage the chatbots' human-like qualities to trick users into sharing sensitive information or clicking on malicious links.

Privacy Concerns: Chatbots collect and store enormous amounts of data, including conversation logs and personal information.



Such stored data is vulnerable to theft and misuse by hackers, which could lead to privacy violations.

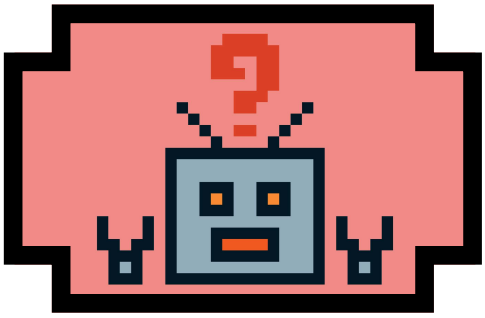
Additionally, persistent chatbot risks remain such as **phishing attacks**, as malicious bots can impersonate legitimate services to steal users' information, requiring caution in verifying the bot's identity before sharing any information. **Malware distribution** is another persistent threat to chatbot use in spreading harmful software. It is advisable to avoid clicking on suspicious links or downloading attachments sent by chatbots. Sensitive, private information shared with bots can be intercepted or misused compromising private data. Every chatbot user should be mindful of **data privacy** disclosure during interactions with AI chatbots.

Chatbot Scams

Advances in AI technology make it harder to identify scams, including chatbot scams. AI hacking is improving by using AI to eliminate spelling and grammatical mistakes, generate email threads with convincing voiceovers, and become more believable and persuasive. Among the effective AI chatbot scams are so-called "hallucinations".

Chatbot "Hallucinations"

So far, AI chatbots and ChatGPT have also provided erroneous answers or outright false information that could be dangerously detrimental, popularly called chatbot "hallucinations." Chatbot hallucinations result from an enormous amount of data fed to AI, far more than it can properly store, which comes out compressed, in the form of incorrect or false information.



A case in point is a Paris-based firm, Nabla. When the company tested a text generator by seeking medical advice for fake patients, the AI-generated ChatGPT-3 answers not only acknowledged but encouraged self-harm.

This case points to the potential hazards of relying on AI for guidance, highlighting the critical importance of ethical considerations in AI development, especially in healthcare. It is amply clear that professionals and general digital users cannot rely on AI for sensitive medical advice where erroneous information can be devastating. The same is true for any professional field where chatbot/ChatGPT information cannot be applied with final, authoritative certainty.

Ways to Spot Fraudulent AI Chats and Stay Safe

While chatbot answers can be valuable, and generally safe, a few telltale signs indicate a chatbot scam is targeting you. When a chatbot opens on a website, click only those chatbots of websites you have navigated to yourself (do not use chatbots on websites you reached by clicking on links in suspicious emails or texts).

If you receive an email with a link to a chatbot, always **verify the "from" address** before clicking on any links. For example, if the sender claims to be from Target, the from address suffix should match Target's web address. Click only on links in texts or emails you were expecting, or from senders you know.

Ignore tempting offers and incredible prizes, especially those appearing out of nowhere. Chatbot scams typically originate from pop-ups and links from websites, emails, and text messages.

Text messages have an advantage over email scams in that shortened URLs hide questionable URLs and abbreviated or incorrect grammar more easily.

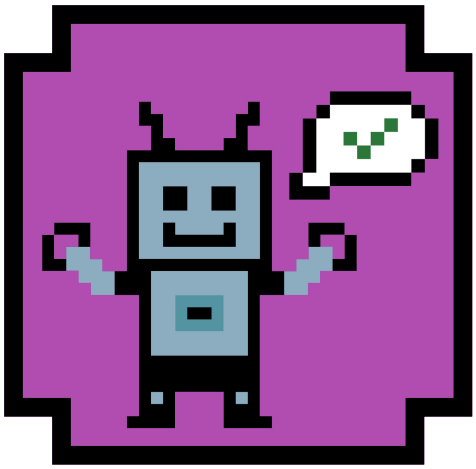
Remember: if an offer seems too good to be true, it probably is!

For all **suspicious messages**, please check out the company name, their offer, and the message. If the offer is real and valid, you will probably find more information on the company's website. If the offer is a scam, you might find news reports about the scam or no information. However, a two-factor or **multi-factor authentication** will further protect your accounts from unauthorized users.

As a rule, random requests for your payment information or **personal details via chatbots** should be ignored. Legitimate companies do not ask personal, sensitive questions via chat. Staying vigilant about suspicious chatbot messages and reporting malicious activity is a way to be proactive against chatbot scams. And if possible, refrain from using unsecured or public Wi-Fi networks to access sensitive information.

Rely on **trusted services** from reputable companies, especially those whose legitimacy you can easily verify. Research and read reviews to ensure the service is trustworthy. Ensure the service has the appropriate security certification or **security seals** from recognized authorities, for an added layer of trust.

Never share **sensitive data**, such as personal, financial, or login details with AI chatbots. If a bot asks for such information, it's a red flag. **Strong passwords**, complex, and unique are a good defense against bot scams.



Software should be regularly **updated**, as regular updates help protect against the latest threats. Chatbots should especially be updated, as regularly updated chatbots are less likely to be open to vulnerabilities.

Stay vigilant about online activities and regularly check statements from the bank or credit card for unauthorized transactions. Early detection can mitigate potential damage. Keep up with the latest **cybersecurity news** to be aware of new threats. Staying informed allows you to take timely actions to protect yourself. Immediately **report suspicious activity** to the service provider if there is any suspicion a bot is malicious. Quick reporting can help prevent further misuse.

NOTE: The links and products in this document are for informational purposes only and do not signify an endorsement.

IHS Updated Password Procedures

The IHS Office of Information Technology (OIT) has recently implemented a security measure to ensure compromised employee passwords do not put the IHS network at risk. The OIT is now checking IHS D1 passwords against a list of known, compromised passwords. If a match is found, the employee will be prompted to change their password to a non-compromised password.

Additionally, as employees are regularly prompted to change or reset their passwords, these new passwords will be compared to a list of compromised passwords. If an employee tries to change or update their password to a compromised password, an error alert, and a prompt to choose a different password will follow. However, this error alert will not have an explanation. The employee will only be prompted to set a different password. For questions, please contact cybersecurity@ihs.gov.