Jackson Hallahan, Ruby Jiang, Abril Beascoechea, Day Wade, and Lausdrith Garcia  
CIS 483  
Prof. Sadaf Ahmed  
10 December 2024

Mitigating AI-Driven Risks in Database Security: Challenges and Actionable Solutions

# Abstract

Recent years have been marked with rapid advancements in Artificial Intelligence (AI) and Machine Learning (ML). As new use cases for these technologies continue to develop and emerge, AI has become a common subject of debate due to its high potential and relatively indeterminate risk. Among the numerous risks posed by AI, three examples of threats posed to database security are the uses of chatbot-assisted SQL Injection, AI-assisted password cracking, and AI-driven impersonation methods. While traditional defensive measures, such as input validation, encryption, strong password policies, and multi-factor authentication (MFA) should all be rigorously updated to account for newer and more advanced AI-assisted penetration attempts, it is also more important than ever that administrative personnel be regularly retrained and kept up-to-date on emergent technologies and potential AI-driven threats that can be stopped with human diligence and intervention. Emergent technologies such as artificial intelligence can also be leveraged for defensive purposes as well, as they can be leveraged to create more advanced detection for SQL injections and AI-driven impersonation.

Ultimately, it is recommended that database developers and administrators take proactive and aggressive precautions to prevent current vulnerabilities from being exploited while mitigating future risks as the technology develops in popularity, capability, and complexity.

# AI-Driven Risks to Database Security

## Introduction

Recent years have been marked with rapid advancements in Artificial Intelligence (AI) and Machine Learning (ML). As new use cases for these technologies continue to develop and emerge, particularly through the leverage of generative tools such as ChatGPT and Bard, AI has become a common subject of debate due to its high potential and relatively indeterminate risk regarding its growing popularity and unregulated use due to potential security, data integrity, and privacy concerns (U.S. Government Accountability Office). The lack of policies governing the use of AI technologies, particularly regarding database security and access control, means that proactive measures must be taken to protect databases against potential new threats brought on by advancements in AI technology.

This paper seeks to address the potential threats that may compromise the security and integrity of database servers, as well as present possible preventative measures and policies that could mitigate these risks. Among the numerous risks posed by the use of AI, this report explores the threats posed by chatbot-assisted SQL Injection, AI-assisted password cracking, and AI-driven impersonation methods.

## Threat 1: Exploiting AI chatbots through SQL Injection

SQL injection attacks are one of the most common web vulnerabilities, currently ranking third in the Open Web Application Security Project (OWASP) Top 10 most critical security risks to web applications. With their “high-impact severity” potential, these attacks are further complicated with the integration of Language Learning Models across all industries online. The widespread adoption of chatbots for customer interaction has introduced a new potential attack vector. Not only should developers and database administrators focus on threats posed by traditional web app injection attacks in input fields but should consider the new potential entry points in systems that have integrated chatbots.

Chatbots and their backend integration can be exploited in many ways. Potential exploits include unauthorized data retrieval, data manipulation, service disruption, all of which can negatively impact businesses and damage their credibility. “An attacker could create a direct prompt injection for the language model being used, telling it to disregard the system prompts set by the application’s creator” (Morris, 2024). Once safeguards are bypassed, attacks can leverage the bot to execute malicious code prompts, extract sensitive information by way of social engineering, and even deploy spam or denial-of-service attacks to other users. Not only are businesses at risk, but users who interact with compromised chatbot “can result in falling victim to phishing scams, system compromises or disclosing personal information” (Morris, 2024).

While SQL injection remains one of the easiest deployed attacks, poor database configuration and input sanitization practice will allow attacks to continue. Exploiting chatbots has only added a new exploit to the issue. Without considering increased security measures in both chatbot systems and web applications, these risks will persist.

### AI-Assisted SQL Injection: Preventative Measures and Policies

While the integration of chatbots can increase the likelihood of falling victim to injection attacks, there are a few preventative measures that should be practiced. Instead of just avoiding the use of LLM bots altogether, organizations can reduce injection attacks by validating inputs, closely monitoring bot activity, and keeping users in the loop (Kosinski, 2024).

Chatbots can potentially mistake malicious code as common human language, making input validation and sanitization of user’s queries is an ideal first line of defense. If bots are susceptible to running malicious queries, signature-based filters should be set to validate proper user input length, and flag any syntax commonly used in injection attacks. Additionally, developers can leverage artificial intelligence to train machine learning models to act as a first line of detection (Kosinski). Parameterized queries are another effective way to ensure user input isn’t mistaken as executable code, and simply treated as data (Amah, 2024). Input validation and parameterization are effective measures for traditional web applications, but researchers at UC Berkely have proposed “structured queries” as a solution that converts user input into a format that is readable by LLM bots. “Our system significantly improves resistance to prompt injection attacks, with little or no impact on utility” (Chen et al, 2024).

For robust policy implementation, securing the database can be done through access control, human approval, and chatbot output filtering. Controls such as least privilege can help ensure chatbots have limited access to the database and operate with minimal permissions to minimize the attack surface. Developers can also integrate an element of human approval before bots can output sensitive data or change database settings (Kosinski). At the data layer, data in transit should be encrypted using Transport Level Security (TLS) and Transparent Data Encryption (TDE) can secure data at rest. Combining the preventative measures and policies discussed with proper security training for administrators, developers, and end users will allow chatbot integration with a reduced risk of SQL injection vulnerability.

## Threat 2: AI-Assisted Password Cracking

AI-assisted password cracking poses a significant and escalating threat to database security, leveraging advancements in artificial intelligence to enhance the speed and sophistication of traditional attack methods. Tools powered by deep learning models, such as PassGAN (Password Generative Adversarial Network), represent a new frontier in password-cracking capabilities. These tools are capable of rapidly analyzing patterns in commonly used passwords, identifying weak points, and generating millions of variations with unprecedented efficiency. PassGAN, for example, has demonstrated the ability to crack 51% of common passwords in under a minute, underscoring the alarming capabilities of these technologies (Tom's Hardware). This development represents a shift in the cybersecurity landscape, where attackers no longer rely solely on manual or traditional automated methods but instead exploit the computational power and adaptability of AI systems. This rapid advancement has created an urgent need for robust and adaptive security measures to protect sensitive data and prevent breaches.

### AI-Assisted Password Cracking: Preventative Measures and Policies

To address this growing threat, organizations must adopt a multi-faceted approach to security. Transitioning to passwordless authentication methods—such as biometric systems, hardware tokens, or multi-factor authentication (MFA)—can drastically reduce reliance on traditional passwords, thereby eliminating a common vulnerability. Biometric authentication, for instance, ties access to unique physical characteristics, making it nearly impossible to replicate. In parallel, leveraging AI-powered countermeasures to detect and block suspicious activities can significantly enhance security. Machine learning algorithms can identify anomalies in login patterns, flagging potential automated password-cracking attempts in real-time and reducing response times to active threats (EC-Council).

Strengthening password policies also remains a critical component of the defense strategy. Policies that require users to create lengthy, complex passwords—preferably exceeding 18 characters and incorporating symbols, numbers, and mixed-case letters—add an additional layer of protection. Combined with measures such as rate-limiting login attempts, these policies can effectively deter even the most advanced AI-driven cracking methods. Research indicates that well-crafted passwords, especially those created using password management tools, are far more resistant to AI-based attacks (Cybernews).

Education and training are equally crucial in combating the risks posed by AI-assisted password cracking. Users and administrators must be equipped with the knowledge to recognize phishing attempts, secure their accounts, and understand the consequences of using weak passwords. In addition, encrypting and salting stored passwords adds another layer of protection, ensuring that even if credentials are stolen, they remain unreadable without significant computational effort. Encryption transforms passwords into unreadable strings, while salting introduces unique, random data to each password, rendering stolen data significantly less useful to attackers.

Combating the threat of AI-assisted password cracking requires more than a static set of solutions; it demands continuous monitoring, adaptive strategies, and regular updates to security protocols. As AI technology continues to evolve, so too must the measures used to safeguard databases. Only by implementing these comprehensive strategies can organizations effectively mitigate the risks posed by these sophisticated and rapidly advancing threats.

## Threat 3: AI-Driven Impersonation

Artificial Intelligence is rapidly advancing, and a new and alarming form of scamming known as AI-driven impersonation, has emerged. AI-driven impersonation leverages advanced machine learning models to convincingly simulate individuals or automated systems. Meaning, it has the ability to generate highly realistic audio and visual content. "From voice cloning to deepfake technology, scammers now possess tools capable of deceiving even the most discerning individuals. The enhanced realism of these impersonations makes it increasingly challenging for targets to distinguish between genuine and fake content" (BrandShield). This threat exploits vulnerabilities in identity verification processes, leading to unauthorized access, data breaches, and manipulation of sensitive information.

Deepfake technology is a prominent example, enabling attackers to create hyper-realistic voice, video, or text simulations of trusted individuals. Such impersonations can trick employees into sharing confidential information, approving unauthorized transactions, or granting access to restricted systems. Attackers may also use AI to mimic customer service representatives or technical support staff, deceiving users into revealing sensitive data or login credentials.

AI-generated impersonations can also be used to mimic the voices or appearances of public figures. Politicians, celebrities, and influencers run the risk of scammers exploiting their recognition to deceive the public, spread false information, or engage in fraud. The risks could have significant reputation and financial consequences.

AI-powered impersonation is not limited to individual targets. It extends to large-scale phishing campaigns, where machine learning algorithms can craft highly personalized messages by analyzing massive amounts of data about potential targets from social media, email, and other online interactions. These messages appear authentic, increasing the likelihood of users falling victim to scams or malware. They can even create entire websites that resemble legitimate ones to trick users into revealing their information.

### AI-Driven Impersonation: Preventative Measures and Policies

Scammers are likely going to leverage the capabilities of AI-impersonation to create even more sophisticated and convincing scams. To protect ourselves and mitigate this threat, we must adopt a multi-layered approach. Organizations should implement robust identity verification mechanisms, such as multi-factor authentication and biometric verification, to prevent unauthorized access. At the same time, there should be an implementation of advanced detection technology that uses AI itself as AI-driven anomaly detection systems that monitor irregularities in user behavior or communication patterns, to flag potential impersonation attempts. Furthermore, regular training programs should be conducted to educate employees about recognizing signs of potential scams, pay attention to possible social engineering tactics, and respond appropriately.

By combining advanced technological safeguards with adapted security and proactive awareness measures, organizations can minimize the risks associated with AI-driven impersonation and protect their information and clients from this sophisticated threat.

## Conclusion

As AI continues to evolve, it is vital that administrators and developers remain diligent in updating policies to account for new threats posed by AI-assisted penetration. It is more crucial than ever to keep users and database personnel up to date while making sure to refresh defensive measures and policies such as input validation, encryption, password policies, and MFA to mitigate AI-empowered threats such as SQL injections, password cracking, or impersonation. Additionally, just as attackers may utilize emergent technologies such as AI to enhance their penetration methods, developers could also leverage these technologies to empower defensive measures, such as for advanced detection against AI-powered injections and impersonation. Ultimately, it is recommended that developers and administrators take proactive and aggressive precautions to prevent current vulnerabilities from being exploited while mitigating future risks as the technology develops in popularity, capability, and complexity.

# References

Amah, D. (2024, July 1). *Understanding SQL injection vulnerabilities in LLM Chatbots: Chatgpt, claude, and Gemini.* Codementor. <https://www.codementor.io/@danielamah/understanding-sql-injection-vulnerabilities-in-llm-chatbots-chatgpt-claude-and-gemini-2hs2lmhiit>

BrandShield Blog. *Ai impersonations: Unveiling Scamming’s new frontier*. Retrieved from <https://blog.brandshield.com/ai-impersonations-unveiling-scammings-new-frontier>

Chen, S., Piet, J., Sitawarin, C., & Wagner, D. (2024). StruQ: Defending Against Prompt Injection with Structured Queries. *ArXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2402.06363>

Cybernews. (2023, April 10). AI can crack half of common passwords in less than a minute. Retrieved from <https://cybernews.com/security/ai-password-cracker/>

EC-Council. (2023, December 2). Role of AI & ML in Enhancing Cybersecurity Against Threats. Retrieved from <https://www.eccouncil.org/cybersecurity-exchange/network-security/role-of-ai-ml-in-enhancing-cybersecurity-against-threats/>

Home Security Heroes. (2023, April 9). AI-Powered Password Cracking Tools. Retrieved from <https://www.gizmochina.com/2023/04/09/ai-powered-password-cracking-tools/>

‌Kosinski, M. (2024, November 7). *Protect against prompt injection*. IBM. <https://www.ibm.com/think/insights/prevent-prompt-injection>

‌Morris, C. (2024, July 23). *Exploring chat injection attacks in AI systems*. Prism Infosec. <https://prisminfosec.com/exploring-chat-injection-attacks-in-ai-systems/>

OWASP Foundation. (2021). *Owasp Top Ten*. OWASP Top Ten | OWASP Foundation. <https://owasp.org/www-project-top-ten/>

TechCircle. (2023, April 10). AI might crack over 50% of passwords in less than a minute: Study. Retrieved from <https://www.techcircle.in/2023/04/10/ai-might-crack-over-50-of-password-in-less-than-a-minute-study>

Tom's Hardware. (2023, April 10). AI Can Crack Most Common Passwords In Less Than A Minute. Retrieved from <https://www.tomshardware.com/news/ai-cracks-most-common-passwords-in-less-than-a-minute>

Treml, F. (2022, January 21). *Top 3 chatbot security vulnerabilities in 2022*. dzone.com. <https://dzone.com/articles/top-3-chatbot-security-vulnerabilities-in-2022>

U.S. Government Accountability Office. (2023, September 6). *Science & Tech spotlight: Generative AI*. Science & Tech Spotlight: Generative AI. <https://www.gao.gov/products/gao-23-106782>