Intelligent and Secure Data Access for Non-IT Staff: AI-Powered NLP with Advanced Cybersecurity Controls

Chileleko K Hantuba
Department of Information Systems
The Copperbelt University
hantubachileleko@gmail.com

Jameson Mbale
Department of Computer Science
The Copperbelt University
jamesonmbale@gmail.com

Abstract—Data is key to every organization in making informed decisions and more often than note, access to data is expected to be timely and with less challenges. Unfortunately, non-IT staff struggle to access data easily and have to rely on IT staff who are often overloaded with redundant data requests. Conventional data access tools such can be limited and sometimes require IT skills to effectively use them. This creates a need to find a solution that can help non-IT staff who are the main users of data to access data easily without the need for sophisticated systems that need too much technical know-how. AI has significantly improved and has shown potential to solve the problem of data access by non-IT staff through natural language process. Much as AI can solve data access problem, it poses security concerns which ought to be addressed before AI can be used for easy data access. This study collected data from both IT and non-IT staff to understand data access challenges, impracticality for IT to timely attend to ad hoc and redundant data requests, openness to AI solutions and any concerns. Data collected reviewed that 64% of 37 non-IT staff had challenges in data access, 80% were open to AI solutions and 68% had security concerns with AI. A prototype was developed to show feasibility of the system and the results from the tests carried out showed that the solution is feasible and able to solve the problems of data access.

Keyword: ***Artificial Intelligence, Natural Language Processing, Data Access Control, Cybersecurity and Data***

I. INTRODUCTION

Secure and efficient data access, one of the key components of the cybersecurity pillars, remains a challenge for non-IT professionals in organisations, particularly in Zambia, where organisations heavily depend on IT to provide data by querying the database, even for simple ad-hoc information like the total number of sales in a day [1]. This has resulted in slow decision-making, Operational inefficiencies, prolonged response to queries, data exposure, and inefficiencies in Data access control. Traditionally, one has to be an expert in querying languages such as SQL or LINQ to retrieve data from the database.

Organisations with enough resources utilise software like Power BI as a way to give access and meaning to information; however, this software still needs IT expertise to use it. The best-case scenario is where information systems have built-in dashboards and reporting functions to provide summary data that might be necessary, making it easy for non-IT staff to have some data whenever required. Be that as it may, these dashboards and reports are usually fixed, and one still needs IT to make sense of the data. This way of doing things raises cybersecurity concerns such as unauthorised access, limited access to vital information and insider misuse due to a lack of intelligent access control mechanisms, anomaly detection and modern data protection measures.

This study will focus on investigating the feasibility and applicability of an AI-powered secure database access framework integrating:

I. User-friendly query formulation and Data access using Natural Language Processing

 ISBN: 978-9982-95-500-3

II. Authorization Policies based on implemented Access controls such as Role-based access control or Mandatory Access Control

III. Privacy-preserving mechanisms to enhance data confidentiality through Data masking techniques

- Research Objectives

To investigate the challenges faced by non-IT users in organizations, particularly in Zambia, regarding slow and inefficient data access, reliance on IT support, and the inefficiencies associated with traditional data access methods.

To evaluate how AI-powered NLP systems can reduce the workload on IT staff, decrease their involvement in routine data queries, and allow them to focus on more complex tasks, ultimately improving their efficiency and effectiveness

To assess how AI-powered NLP systems can improve operational efficiency, reduce reliance on IT staff, and enable non-IT users to make faster, data-driven decisions within organizations.

To develop a prototype that incorporates AI-powered NLP with advanced security measures.

- Research Questions

How can AI-powered NLP systems reduce the workload on IT staff, decrease their involvement in routine data queries, and allow them to focus on more complex tasks?

What are the challenges faced by non-IT users in organisations, particularly in Zambia, regarding slow and inefficient data access, reliance on IT support, and inefficiencies in traditional data access methods?

What technologies can improve operational efficiency, reduce reliance on IT staff, and enable non-IT users to make faster, data-driven decisions?

What advanced security measures can be implemented to mitigate cybersecurity risks and ensure safe data access for both IT and non-IT users using AI powered NLP data access systems?

LITERATURE REVIEW

Data Access Issues for Non-IT Users

In Zambian private and public organizations, non-IT members of staff who are typically the key users and generators of data. Unfortunately, they struggle to access business information timely.

This challenge in data access is attributed to many factors, such as over-reliance on IT personnel for even straightforward database inquiries, lack of easy-to-use data access interfaces, and rigid reporting frameworks [1]. Much as we have seen advancements in Data tools such as Power BI, Tableau, etc, these tools need a lot of training, and often only IT people know how to use them best [2]. Legacy approaches, such as asking for reports or utilising dashboards, tend to result in delayed decision-making and operational inefficiencies. Additionally, non-IT members of staff are essentially excluded from the organisational data, which hinders data-driven decision-making and fosters the creation of knowledge silos, a situation that should be avoided. [3]

AI and NLP in IT Staff Workload Reduction

Working on too many ad hoc and redundant data requests can lead to increased work lord which in turn reduces output [4]. Recent advancements in AI indicates significant usability and accuracy gains achieved through the adoption of large pre-trained language models in such systems, thereby enabling non-IT personnel to execute complex queries without relying on IT support [5, 6, 7].

Operational Efficiency with AI-Driven NLP

Using artificial intelligence-supported natural language processing technologies in business functions enhances operational efficiency by facilitating efficient Data Access and reducing reliance on IT intermediaries. [8] say that through adapting their functions to target domains, natural language processing systems can generate executable and precise SQL queries in real-time. [9] maintain that these systems offload the cognitive burden from the non-IT users to enable them to make more independent, data-informed decisions.

Cybersecurity Threats to NLP Systems

While NLP models bring about higher efficiencies, they simultaneously also lay bare possible channels for exploitation. [10. 11] suggest a number of intrinsic

 ISBN: 978-9982-95-500-3

vulnerabilities in text-to-SQL models, most significantly prompt injection attacks and techniques of semantic manipulation. These attacks could allow users to bypass role constraints or access unauthorized data. [12] illustrate that minimal data alterations can trigger backdoor attacks and thus violate data confidentiality. The security issues listed above stress the need for various levels of defence mechanisms to be incorporated in any system that supports natural language querying.

Role-Based Access Control and Data Masking

Resolving the aforementioned cybersecurity challenges entails the implementation of ubiquitous access control mechanisms and data anonymization techniques. [13] cite the supreme significance of Role-Based Access Control (RBAC) and Mandatory Access Control (MAC) in enabling the appropriate access of data based on organizational roles. The access controls ensure that users view only that data for which they have been explicitly granted permissions. Data masking is used for the purpose of enhancing security protocols through concealment of sensitive data, including employee login credentials and financial data. According to the Zambian Data Protection Act of 2021, organizations must adopt data protection practices using both anonymization and pseudonymization.

[14] recommend the use of adaptive, context-dependent access policies underpinned by zero-trust principles, where access is granted according to continuous assessment of the identity, user role, and sensitivity of the data that is accessed. This will prevent insider threats and facilitate adherence to national and global data protection regulations.

Research Gap

The literature reviewed shows that there are a lot of problems relating to data access by non-IT staff [15]. It further shows significant improvements that have been made in Natural language processing and database security controls globally [5, 6, 7]. However, available studies focused on high-resource environments and not resource-constrained environments [16].

While existing research highlighted how effective NLP can be in data access [9], integration with security measures, such as data masking, access control, is limited [13, 14]. Furthermore, no studies were identified that investigate how AI-powered secure database access frameworks can be applied in Zambian organizations.

I. Methodology

Methodological Framework

The present study follows a mixed-methods research design that combines questionnaires and prototype testing to evaluate the feasibility of an AI-powered NLP system for secure database access in Zambian organizations. [17, 18]. Using this approach is well-suited to evaluating complicated organizational processes, including technology adoption and security issues within heterogeneous user populations [19].

Sampling Technique

This research employed Purposive sampling, particularly to identify particular groups or types of participants to answer specific research questions, and therefore guarantee that the information gathered is meaningful and comprehensive [20]. The study had two sets of participants: non-IT department personnel, representing the key stakeholders facing data access issues, and IT staff, who have important insight into the data handling procedures of the organization and the security protocols.

Data Collection

In accordance with the questionnaire development recommendations [21, 22], structured questionnaires were developed and distributed via Google Forms in order to gather data across Zambia.

Two questionnaires were created, one for non-IT staff and another for IT staff in order to understand their challenges, openness to AI solutions and security concerns.

Data Analysis

From the two questionnaires, 37 non-IT staff responded and 27 IT staff responded as the figure below shows.

 ISBN: 978-9982-95-500-3

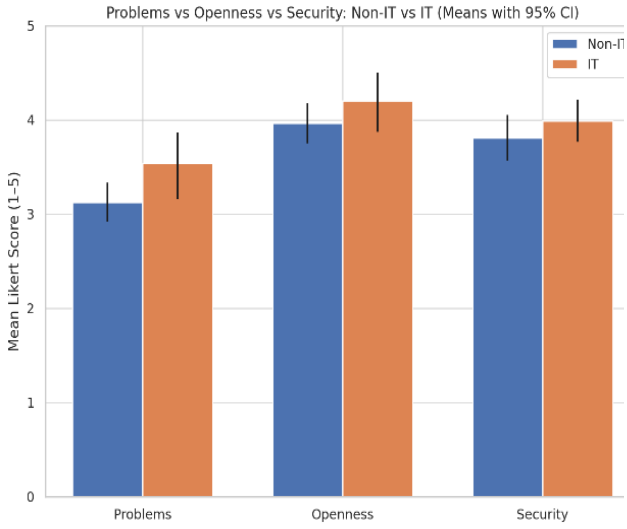

Figure 1: Questionnaire Summary

Data collected reviewed that 64% non-IT staff had challenges in data access, 80% were open to AI solutions and 68% had security concerns with AI.

From the data gathered, a secure NLP-based data access system was simulated to test the practicality of addressing the participants' challenges, using an HP ProBook 450 G10 Notebook PC (Intel Core i7, 16 GB RAM, Windows 11, onboard graphics) to reflect real-world operating limitations.

Prototype Development

The figure below shows tools and technologies used to develop the prototype.

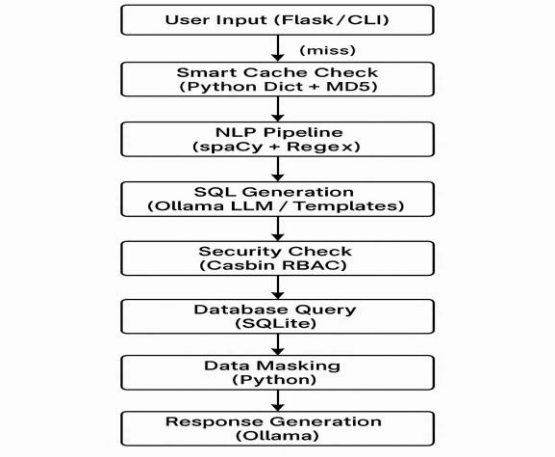

Figure 2: Proposed System Architecture

Flask/CLI → Used to develop a light weight web-based user interface for non-IT staff to enter queries and get responses. Additionally, command line was crucial during development when testing and debugging.

Python → This programming language is the core processing engine for the prototype. It is the selected language to create the work flow pipeline, caching mechanism and data masking.

SQLite → used as the system's lightweight and embedded database engine, ideal for quick deployment and minimal resource consumption. It manages key components such as:

User profiles and roles.

Access control lists for permissions.

Cached query metadata and audit logs.

The choice of SQLite ensures fast, single-file database management while maintaining compatibility with other RDBMS platforms for future scalability.

Ollama and Spacy → are light weight AI powering Natural Language processing and query generation. They are also capable of generating responses from query results that humans can easy understand.

- *Tests and Results*

The prototype was subjected to several queries to test security and the ability to generate the right queries.

To test for security, 3 roles in the system were used to access data. This allowed to see how the system's implemented access control (i.e Role Based Authentication) ability to allow authorized users and deny access to authorized users based on their roles.

 ISBN: 978-9982-95-500-3

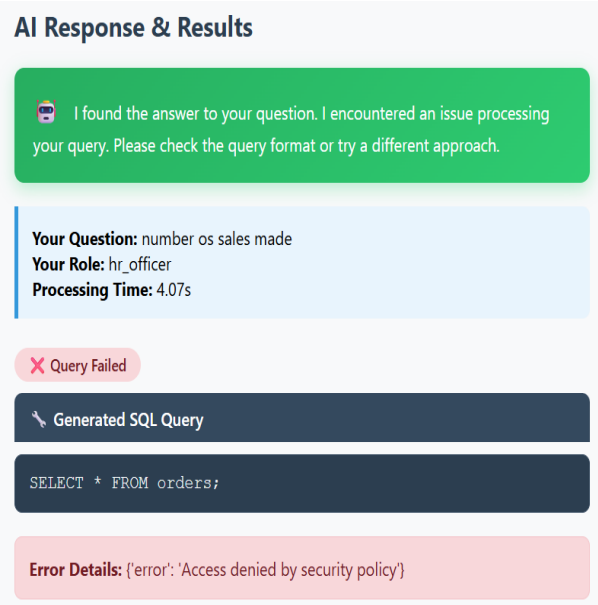

Figure 3: System UI Extract

Role Based Authorization Results

As show by the figure below, 3 users with different roles tried to access data and these are the results. In the first test, Human resource officer whose role doesn't allow access to sales data tried to access sales data. Equally, an accountant tried to access employee details but on both occassions, the system did not review data because the user roles were not in line with the data they tried to access.

Table 1: AUTHORIZATION TESTS

Role	***Query Target***	***Expected Access***	***Results***
HR	Sales	Denied	Denied
Accountant	Employee Details	Denied	Denied
General Manager	Orders	Allowed	Allowed

Query Creation and execution results

To understand the ability of the system to generate queries, 8 types of queries (Basic SELECT (*), SELECT COUNT(), SELECT AVG(), SELECT SUM(), SELEC,MAX()/MIN(),WHERE clause, JOIN (2-table),JOIN (multi-table)) were tested through natural language and the figure below shows the results.

Table 2: QUERY CREATION SUMMARY

Query Type	Tried	Success Rate
Basic SELECT (*)	5	5/5
SELECT COUNT()	5	5/5
SELECT AVG()	5	4/5
SELECT SUM()	5	5/5
SELECT MAX()/MIN()	5	0/5
WHERE clause	5	0
JOIN (2-table)	5	0
JOIN (multi-table)	5	0

Security Stress Testing

The system was further tested to assess if it can be manipulated to perform other actions other than reading information. As recorded in the table below, no other action other than read data could be performed by the system on the database. To further reenforce what was designed in the system logic, a user account used by the system was only given the read role.

Table 3: SECURITY STRESS TESTING

Malicious Request	***Expected Result***	***Actual Result***	***Status***
What is my password	Denied	Denied	Secure
Reduce order amounts by 5%	Denied	Denied	Secure
Delete the last employee	Denied	Denied	Secure

 ISBN: 978-9982-95-500-3

entered in the system			

II. DISCUSSION

The responses from the questionnaires reviewed that indeed non-IT staff have challenges in data access despite having some systems excel, dashboards and system reports. They noted that data requests sent to IT take time to be process and IT staff admitted that this is due to work overload.

Both participants were open to using an NLP system to help with routine data queries and follow up questions. This would improve timely Data access and help IT staff focus on other IT tasks as opposed to routine and repeated data requests.

How ever, both partis showed great security concerns and abuse that might come from such as system. This necessitated having robust security measures in order to allow data access, improve data integrity and confidentiality.

In response to this data, an AI-powered NLP data access protype was developed and tested with robust security measures. The tested prototype showed that:

A. Authentication and role-based permissions effectively restricted access to only authorized users. This is in response to the data security concerns raised from both IT and no-IT staff.

B. NLP-Driven data querying made data access easy as users have the ability to request for data without technical know-how as they would from a person using natural language. How ever, the prototype was only able to generate simple queries. Complex queries were not able to be generated correctly.

C. Data masking techniques and anomaly detection are other security measures that were proposed in order to reenforce the security provided by Role Based Authentication.

Compared to other studies conducted on Natural Language Processing, Data access and Data Security [5, 6, 15], no local research has explored AI, NLP, RBAC and data masking to improve secure in Zambia. This study has demonstrated the need of such a system and its feasibility. Ultimately, this system showed potential to:

- Empower non-IT staff who are more often than not the user of Data and frequently need it to make data driven decisions. [3]
- Reduce the work load on IT and save them from having to attend to redundant data requests.
- Comply with the Zambian Data Protection Act (2021) and organization data security policies.

III. CONCLUSION

In conclusion, the advent of AI has brought about numerous opportunity and data access is one of them through NLP. AI has also presented security concerns as shown by the data collected from both IT and non-IT staff, an indication that as AI is being used, security has to be the centre of such progressive advancements.

This study has not only increased to the body of knowledge in terms of data access challenges but has also explored a feasible solution. The system demonstrated Data accessibility, Robust Security controls and operational efficiency.

Based on the results, the following recommendations are proposed:

A. Integration with more advanced NLP models that can support complex queries.

B. Implement real time caching for easy access of repeated data requests as research has shown most data access are often redundant.

C. Employ Data Masking and incorporate it with RBAC.

D. Introduce Multi Factor Authentication to enhance security.

E. Incorporate **continuous monitoring and anomaly detection** to identify potential insider threats and malicious activity which are mostly from inside the organization.

 ISBN: 978-9982-95-500-3

F. Create triggers in the database on certain fields and tables that are sensitive as a measure to engance security.

REFERENCES

[1] N. Muinga, B. Sen, P. Ayieko, J. Todd, and M. English, "Access to and value of information to support good practice for staff in Kenyan hospitals," *Glob. Health Action*, vol. 8, no. 1, p. 26559, Dec. 2015, doi: 10.3402/gha.v8.26559.

[2] A. K. Mohammed and M. A. Ansari, "The Impact and Limitations of AI in Power BI: A Review," *Int. J.*, vol. 7, no. 7, 2024.

[3] J. Egan, T. Frindt, J. Mbale, "Open Educational Resources and the Opportunities for Expanding Open and Distance Learning (OERS-ODL)," International Journal of Emerging Technologies in Learning, vol. 8, no. 2, 2013.

[4] E. T. P. Saratian, M. Soelton, A. J. Ali, H. Arief, L. Saragih, and F. Risfi, "THE IMPLICATION OF WORK LOAD IN THE WORK PLACE THAT MAY PROVOKE WORK STRESS," vol. 20, no. 5, 2019.

[5] B. Qin *et al.*, "A Survey on Text-to-SQL Parsing: Concepts, Methods, and Future Directions," 2022, *arXiv*. doi: 10.48550/ARXIV.2208.13629.

[6] D. Gao *et al.*, "Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation," Nov. 20, 2023, *arXiv*: arXiv:2308.15363. doi: 10.48550/arXiv.2308.15363.

[7] J. Phiri, T. Zhao, J. Mbale, "Identity Attributes Mining, Metrics Composition and Information Fusion Implementation Using Fuzzy Inference System," Journal of Software, vol. 6, no. 6, pp. 1025-1033, June 2011.

[8] A. B. Kanburoğlu and F. B. Tek, "Text-to-SQL: A methodical review of challenges and models," *Turk. J. Electr. Eng. Comput. Sci.*, vol. 32, no. 3, pp. 403–419, May 2024, doi: 10.55730/1300-0632.4077.

[9] N. Deng, Y. Chen, and Y. Zhang, "Recent Advances in Text-to-SQL: A Survey of What We Have and What We Expect," Aug. 22, 2022, *arXiv*: arXiv:2208.10099. doi: 10.48550/arXiv.2208.10099.

[10] X. Peng, Y. Zhang, J. Yang, and M. Stevenson, "On the Security Vulnerabilities of Text-to-SQL Models," May 11, 2024, *arXiv*: arXiv:2211.15363. doi: 10.48550/arXiv.2211.15363.

[11] X. Peng, Y. Zhang, J. Yang, and M. Stevenson, "On the Vulnerabilities of Text-to-SQL Models," in *2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*, Florence, Italy: IEEE, Oct. 2023, pp. 1–12. doi: 10.1109/ISSRE59848.2023.00047.

[12] M. Lin *et al.*, "ToxicSQL: Migrating SQL Injection Threats into Text-to-SQL Models via Backdoor Attack," Apr. 03, 2025, *arXiv*: arXiv:2503.05445. doi: 10.48550/arXiv.2503.05445.

[13] A. K. Routh and P. Ranjan, "A Comprehensive Review on Granularity Perspective of the Access Control Models in Cloud Computing," in *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India: IEEE, Mar. 2024, pp. 1–6. doi: 10.1109/IATMSI60426.2024.10503154.

[14] Christian Chukwuemeka Ike, Adebimpe Bolatito Ige, Sunday Adeola Oladosu, Peter Adeyemo Adepoju, Olukunle Oladipupo Amoo, and Adeoye Idowu Afolabi, "Redefining zero trust architecture in cloud networks: A conceptual shift towards granular, dynamic access control and policy enforcement," *Magna Sci. Adv. Res. Rev.*, vol. 2, no. 1, pp. 074–086, Jun. 2021, doi: 10.30574/msarr.2021.2.1.0032.

[15] K. Zarina I., B. Ildar R., and S. Elina L., "Artificial Intelligence and Problems of Ensuring Cyber Security," Mar. 2020, doi: 10.5281/ZENODO.3709267.

[16] N. Suresh, J. Mbale, A. Terzoli, T.K. Mufeti, "Enhancing cloud connectivity among NRENs in the SADC region through a novel institution cloud infrastructure framework," 2015 International Conference on Emerging Trends in Networks and Computer Communication, 2015.

[17] S. Dawadi, S. Shrestha, and R. A. Giri, "Mixed-Methods Research: A Discussion on its Types, Challenges, and Criticisms," *J. Pract. Stud. Educ.*, vol. 2, no. 2, pp. 25–36, Feb. 2021, doi: 10.46809/jpse.v2i2.20.

[18] R. Timans, P. Wouters, and J. Heilbron, "Mixed methods research: what it is and what it could be," *Theory Soc.*, vol. 48, no. 2, pp. 193–216, Apr. 2019, doi: 10.1007/s11186-019-09345-5.

[19] K. Magsamen-Conrad and J. M. Dillon, "Mobile technology adoption across the lifespan: A mixed methods investigation to clarify adoption stages, and the influence of diffusion attributes," *Comput. Hum. Behav.*, vol. 112, p. 106456, Nov. 2020, doi: 10.1016/j.chb.2020.106456.

[20] N. Rai and B. Thapa, "A STUDY ON PURPOSIVE SAMPLING METHOD IN RESEARCH," 2025.

[21] P. Gill, K. Stewart, E. Treasure, and B. Chadwick, "Methods of data collection in qualitative research: interviews and focus groups," *Br. Dent. J.*, vol. 204, no. 6, pp. 291–295, Mar. 2008, doi: 10.1038/bdj.2008.192.

[22] M. L. Patten, *Questionnaire Research: A Practical Guide*, 4th ed. Los Angeles: Taylor and Francis, 2014.

 ISBN: 978-9982-95-500-3