ABSTRACT OF CAPSTONE

Joshua C. Frisby

The Graduate School

Morehead State University

March 8, 2022

INSTITUTIONAL RESEARCH-FOCUSED CONVERSATIONAL ARTIFICIAL INTELLIGENCE AGENT

Abstract of Capstone

A capstone submitted in partial fulfillment of the
Requirements for the degree of Doctor of Education in the
Ernst and Sara Lane Volgenau College of Education
At Morehead State University

By

Joshua C. Frisby

West Liberty, Kentucky

Committee Chair: Lenora J. Justice, Assistant Professor

Morehead, Kentucky

March 8, 2022

ABSTRACT OF CAPSTONE

INSTITUTIONAL RESEARCH-FOCUSED CONVERSATIONAL ARTIFICIAL INTELLIGENCE AGENT

Higher education institutions adopt conversational agents, or chatbots, to perform and automate certain business functions. While chatbots exist to support Enrollment Services, Financial Aid, and other departments within an institution, Institutional Research lacks options. Institutional Research supports higher education institutions by providing information and analysis for internal and external stakeholders. This capstone explores the research, design, and development of IRAbot, an institutional research-focused conversational artificial intelligence agent. By leveraging Morehead State University (MSU) data, IRAbot was developed with features that automate and complement standard Institutional Research processes. Implementation of this tool can positively impact MSU's data culture and practices by increasing data availability, streamlining and simplifying data retrieval, and saving Institutional Research resources by cutting down on data requests.

KEYWORDS: Institutional Research, Conversational Artificial Intelligence, chatbots, automation, higher education

____

Candidate Signature

____

Date

INSTITUTIONAL RESEARCH-FOCUSED CONVERSATIONAL ARTIFICIAL INTELLIGENCE AGENT

By

Joshua C. Frisby

Approved by

____
Dr. Christopher D. Howes
Committee Member Date

____
Dr. Timothy L. Simpson
Committee Member Date

____
Dr. Lenora J. Justice
Committee Chair Date

____
Dr. Timothy L. Simpson
Department Chair Date

RULES FOR THE USE OF CAPSTONES

Name	Date

CAPSTONE

Joshua C. Frisby

The Graduate School

Morehead State University

March 8, 2022

INSTITUTIONAL RESEARCH-FOCUSED CONVERSATIONAL ARTIFICIAL INTELLIGENCE AGENT

Capstone

A capstone submitted in partial fulfillment of the
Requirements for the degree of Doctor of Education in the
Ernst and Sara Lane Volgenau College of Education
At Morehead State University

By

Joshua C. Frisby

West Liberty, Kentucky

Committee Chair: Lenora J. Justice, Assistant Professor

Morehead, Kentucky

March 8, 2022

DEDICATION

To Maci.

To my family and friends.

ACKNOWLEDGEMENTS

I am the product of care and support from many. To those who have helped me along the way, thank you for your inspiration and for helping me grow. A special thanks to my wife, Maci, for her unwavering love and encouragement. Thank you so much to my doctoral committee, Dr. Howes, Dr. Justice, Dr. Simpson, and the Morehead State University EdD faculty for their guidance and instruction. I would like to acknowledge the Morehead State University Office of Institutional Research and Analysis members for their help in gathering data for this project. Thanks to the EdTech Platoon cohort members for your support and friendship. I have enjoyed getting to know each of you and am excited to see what the future holds for each of us! Lastly, I want to thank my family and friends for understanding the time commitment required by this curriculum and for offering your support.

TABLE OF CONTENTS

DEFINITION OF TERMS

Artificial Intelligence. Design of intelligent machines and computer programs that enable functionality and foresight within an environment (McCarthy, 2007; Nilsson, 2009).

Chatbot. Software that simulates human conversation using conversational artificial intelligence (Ranoliya et al., 2017).

Data. Digitized information ("Data Definition & meaning," n.d.).

Data analysis. The practice of applying statistical and logical techniques to describe and evaluate data ("Data Analysis," n.d.).

Data availability. Accessibility and readiness of data (Hardwicke et al., 2018).

Data culture. The processes and practices an organization employs to collect, analyze, and deploy data to make better decisions (Díaz et al., 2018).

Database. A collection of stored data ("Database definition & meaning," n.d.).

Family Educational Rights and Privacy Act. A federal law that dictates how educational records are accessed, amended, controlled, and disclosed ("What is FERPA?," n.d.).

Frequency distribution. Representation of the occurrence of values in an interval ("Frequency distribution definition & meaning," n.d.).

Institutional Research. Research conducted at a college or university to provide information and data to promote institutional effectiveness and other initiatives (Saupe, 1990).

Machine Learning. Artificial Intelligence algorithms that improve through experience and the environment (Naqa et al., 2015).

Natural Language Processing. Areas of research and application allowing computers to understand and interpret human language or speech (Chowdhury & Lynch, 1991).

SQL. A language used to store, manipulate, and retrieve data from a database (Groff & Weinberg, 2010).

EXECUTIVE SUMMARY

What is the core of the capstone?

The core of this capstone project is the design and development of a data and institution-focused conversational artificial intelligence (AI) agent, or chatbot. This chatbot, Institutional Research and Analysis bot (IRAbot), highlights ways conversational artificial intelligence agents can be utilized in a higher education setting to improve data availability, empower those seeking data, and lessen the number of data requests. This project aims to demonstrate the value that Institutional Research (IR) offices can garner by implementing a conversational AI agent using Morehead State University (MSU) data and processes. Through the application of chatbots, institutions and their IR departments can enhance the user experience of data-seekers and cut down on resources needed to satisfy them.

MSU is a public university located in Kentucky ("Morehead State University :: About Morehead State University," n.d.). MSU's Office of Institutional Research and Analysis (OIRA) hosts publicly available data through its web pages in the forms of interactive Tableau visualizations and PDF documents. Although data is offered, OIRA receives data requests that could be fulfilled by requestors exploring available self-service options. This project analyzed requests for data MSU's ORIA received in the 2019-2020 academic year and data from the MSU 2019-2020 profile, which fueled the design and development of chatbot functionality.

An institution's data culture reflects the data-informed decision-making processes and behaviors within the organization. The internal and external systems institutions use to generate and exchange data are part of their data economy. The introduction of IRAbot into the MSU data economy can help bolster the institution's data culture and practices. IR departments are primed to deliver deep information-rich interactions by implementing user-friendly conversational AI agents armed with their institution's data. By marrying common trends and phrases used by those requesting data to the MSU OIRA department with IRAbot's conversational AI engine and capabilities, data-seekers can be informed more quickly and efficiently versus traditional methods.

Who is the capstone meant to impact?

IRAbot was designed to impact MSU's OIRA, decision-makers, and data-seekers. OIRA's mission, which includes providing "high-quality information" ("Morehead State University :: Morehead State University Administration," n.d.) to decision-makers, is supported through the application of IRAbot. Coburn and Turner (2012) identify data use as a prominent strategy for improvement in education and note that many federal and state policies and accreditation agencies support or require the practice. Furthermore, they acknowledge collecting data as the first step in a data-informed decision-making approach. Gill, Borden, and Hallgren (2014) developed a conceptual framework for decision-making using data in education, with assembling "high-quality raw data" as the first step. IRAbot assists with gathering information by retrieving data from vetted repositories.

Fontichiaro and Oehrli (2016) state that data can often be hard to comprehend, even in visualized forms. Given that OIRA typically distributes and displays data in many ways, users can be left with inconclusive results if documentation or design are inadequate. These and other issues can cloud their analysis and drive data-seekers to request data from IR offices that exist but are not understandable. The existence of data does not promote availability or accessibility if they are challenging to locate or interpret. IRAbot helps close knowledge gaps and save resources by featuring functionality to answer questions and perform calculations in an ad-hoc fashion without the help of an IR analyst or other internal resources.

With institutions generating and storing more data than ever before (Murray, 2013), the potential for positive data-informed decision-making has never been greater. Murray (2013) states that institutions are underutilizing data to drive systemic change and are instead focusing their efforts on compliance, accountability, and state and federal reporting requirements. By developing IRAbot with institutional data and documentation, data culture and availability can be positively impacted as data-seekers will have near-immediate access to data and clarification through the chatbot interface.

How was this capstone project designed and developed?

Microsoft Bot Framework was the platform used to construct IRAbot. Microsoft Bot Framework is an extensible ecosystem that offers many tools with robust functionality which allows for the development of a full conversational AI experience. The Microsoft Azure cloud hosts IRAbot and allows communication with other services such as Azure Search and Cognitive Services that are utilized to meet design and functionality requirements. IRAbot can be accessed through a web interface, which allows anyone with an Internet-capable device that features a modern web browser to use it and affords the capability for embedding into existing web pages.

To gather data for IRAbot to use, the disaggregated deidentified 2019-2020 MSU profile data was requested and received from OIRA. The MSU profile features "data regarding enrollment, degrees, financial aid and various other components of the students, faculty, and staff at MSU" ("Morehead State University :: Morehead State University," n.d.). The data from OIRA was received in the form of an Excel workbook with four worksheets representing several data stores used for the MSU profile. Not all MSU profile data was present in the Excel workbook as only "directory information" about consenting students could be released from OIRA to comply with the Family Educational Rights and Privacy Act (FERPA). An anonymized student ID was generated upon request to join records across the datasets. The data received was profiled, cleansed, structured, transformed, and validated before exportation from Excel to CSV flat files. The flat files were loaded into Azure SQL using SQL Server Management Studio (see Figure 1) into three tables enrollment, minors, and degrees (see Figure 2).

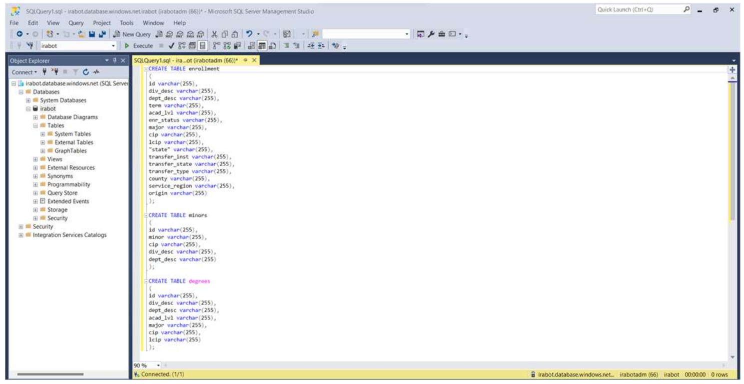

Figure 1. SQL Server Management Studio with IRAbot CREATE TABLE queries.

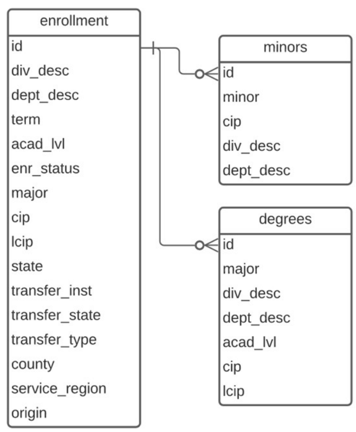

Figure 2. Entity Relationship Diagram (ERD) of database tables.

To define functionality, data requests submitted to the MSU OIRA in the 2019-2020 academic year were also requested and received, then analyzed using the Python programming language and the Natural Language Toolkit (NLTK), which provides natural language processing (NLP) capabilities. The NLTK package allowed the contents of user data requests to be parsed and the most common phrases to be identified through frequency distribution. The top frequency groups (see Figure 3) were later translated into the actions that IRAbot features. Three hundred two data requests were submitted to the office during the 2019-2020 academic year. Of these, 10.6% (32) were determined to be requesting information that is in part available through the public OIRA web presence. For these specific requests, the time from submittal to completion mean is 17.5 days, with 8 days for both median and mode, and the range being 295 days.

The 32 identified data requests were parsed with related words grouped together. Headcount, majors, minors, enrollment status, and graduates were the top five groups that appeared in the data requests and were selected to inform the design of the functionality of IRAbot. Twenty-one of the 32 requests requested majors, 8 requested minors, 7 requested graduates, and 24 requested headcounts of students.

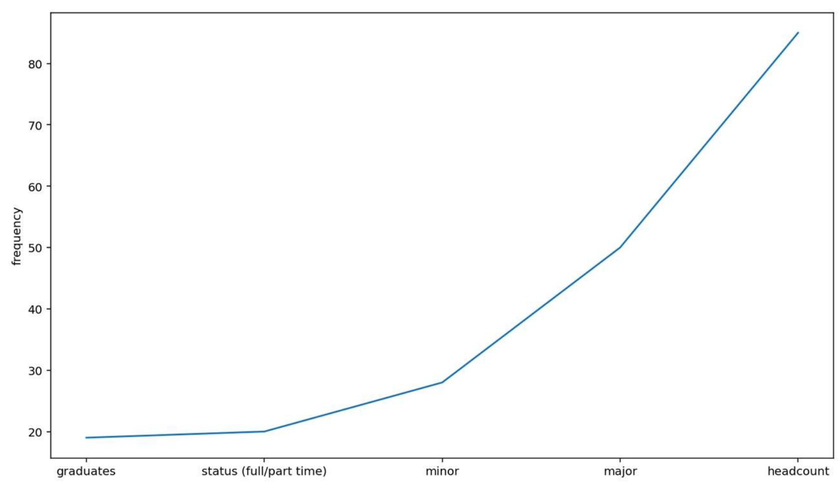

Figure 3. Grouped words frequency distribution.

The development process for IRAbot was based on the Systems Development Lifecycle (SDLC). The stages of SDLC can vary depending on publications and resources but typically feature Planning, Analysis, Design, Development and Testing, Deployment, and Maintenance steps (see Figure 4).

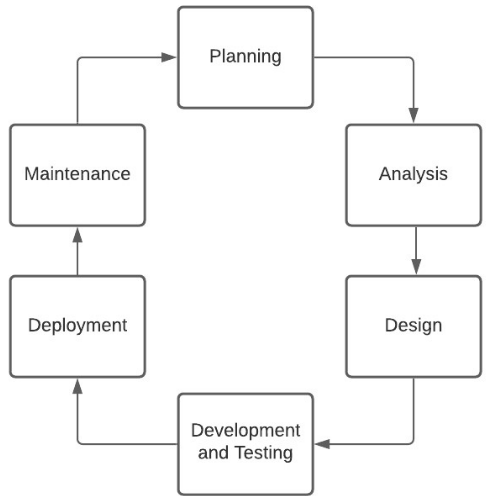

Figure 4. Systems Development Lifecycle.

Planning

Planning a potential solution began with identifying possible enhancements to institutional research practices and offerings through conversational AI. This project was deemed feasible as financial, technical, operational, and schedule needs could be met. The timeline for the project was created and can be seen in the “Plan for Future Implementation” section below.

Analysis

Specifications for a solution with an overall goal of making data easier to access and discover were established. This solution would need to increase data availability and transform the process of data collection for data-seekers and decision-makers. With these considerations in mind, a chatbot platform was determined to be sufficient to meet the project's goals.

Design

Following best practices and research, IRAbot was designed. Specifics related to the design are in the following “Why were this capstone and related strategies selected?” section of this summary.

Development and Testing

IRAbot was constructed with Microsoft Bot Framework and Microsoft Azure services. Microsoft Bot Framework is an extensible ecosystem designed to build conversational AI experiences (“Microsoft Bot Framework,” n.d.). Microsoft Azure, a cloud technology platform, hosts all services related to IRAbot. Users can visit and converse with IRAbot on a web page or in an emulated environment. IRAbot can understand user input through dialog flows and Azure Cognitive Services’ Language Understanding (LUIS) which determines user meaning through AI. QnA Maker is an Azure Cognitive Service that allows the uploading of knowledge base data for use with a chatbot. Azure Search and SQL Database are two other components used to store documents and data for parsing and retrieval. Data received from the MSU OIRA office was used both as a resource for IRAbot to parse for information and as a guide for developing functions and responses. Data was loaded into Azure SQL Database, and SQL queries were developed from Data Request analysis that allows IRAbot to return requested information. Microsoft Bot Framework Composer (see Figure 5) was used to design and build IRAbot flow through a graphical user interface (GUI). Custom actions and functionality were developed within Microsoft Bot Framework Composer and Microsoft Visual Studio using SQL queries, JSON, XML, and the C# programming language. Microsoft Bot Framework Emulator (see Figure 6) was used for local testing.

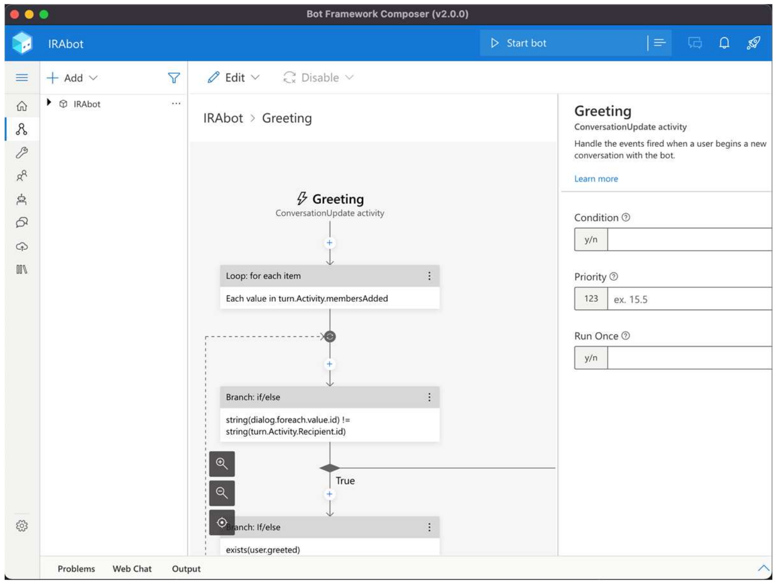

Figure 5. Microsoft Bot Framework Composer.

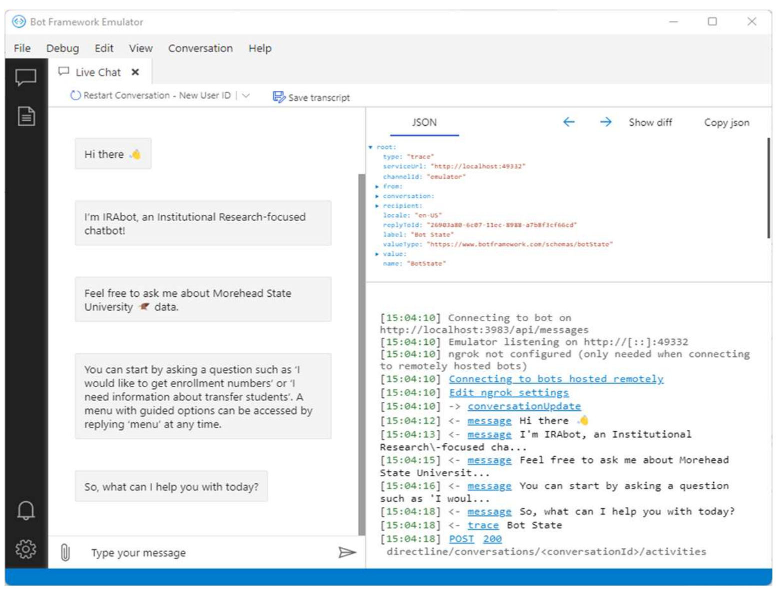

Figure 6. Microsoft Bot Framework Emulator.

The initial architecture for IRAbot leverages Azure Bot Services to communicate with users through a web interface or emulator. Azure Cognitive Services enables IRAbot to interpret user meaning through LUIS and house QnA Maker resources. User queries can be submitted through the IRAbot web interface, and Azure Bot Services will determine the appropriate route to find related information either in Azure Cognitive Services' QnA Maker or Azure Search, which can search documents, information, and data stored in a structured database to return related information to the user. Figure 7 illustrates communication between a user's web client and the IRAbot technology stack.

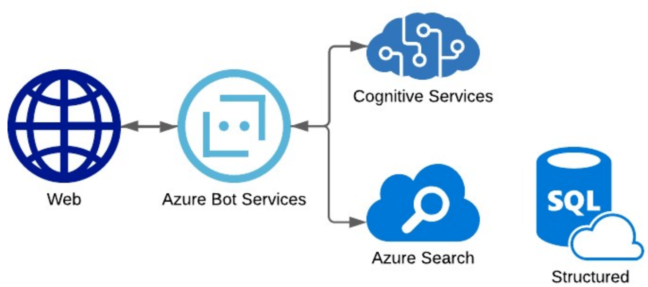

Figure 7. IRAbot hosted technology stack.

The capstone project goals were met when IRAbot sufficiently answered queries and featured required functionality. Development and Testing was the final stage during the capstone timeline. Beyond the capstone, IRAbot will be made available through a web page for a pilot group of testers tasked with retrieving data from IRAbot and noting any issues or flaws during use. Feedback generated by the pilot group tests will be evaluated and considered to improve IRAbot before finally deploying a production-ready build to a public-facing web page. Ongoing maintenance will keep IRAbot, and underlying platforms secure and up to date.

Why were this capstone and related strategies selected?

Institutional Research

The ways Higher Education institutions generate, consume, and distribute data have evolved with advances in technology and data literacy. Historically, institutions have faced “substantial limitations in terms of cost, time requirements, scope, and authenticity” (Greller & Drachsler, 2012, p.42) when dealing with data due to lack of digitization and aging processes. Today, many institutions house rich data repositories and offer publicly available self-service data, typically through their IR or Information Technology (IT) departments. IR’s role in an institution involves providing information and analytics to decision-makers and the campus community and meeting regulatory reporting requirements (Turk and Taylor, 2019). The functions of IR are crucial for institutional success (Letzring, 2017).

MSU’s OIRA meets the university’s internal and external reporting needs using various tools and processes. Working previously in MSU’s OIRA as a Programmer Analyst allowed me to interact with and perform data analysis for several internal and external stakeholders. One service the department offers is a data request form which can be used to solicit data and information about the institution. Although public OIRA departmental webpages offer users a variety of data for consumption, requests are often received for already available data. While data may exist on these web pages, not all users are privy to the processes or technologies these data are presented and hosted on, which can negatively impact data availability and accessibility. One tool that has not yet seen wide adoption by Higher Education IR offices or MSU's OIRA that can address these issues and serve their population's data needs is a chatbot.

Data. Harnessing the power of data is fundamental to identifying trends and planning for the future. Expanding the availability of data aids decision-makers by equipping them with the information needed to pursue action. Data by itself is typically not enough to make decisions. Data can be meaningless without contextualization and analysis (Albright et al., 2011). Educators and institutions must adequately leverage data and experience to inform practices (Mandinach, 2012). Reeves and Burt (2006) explain that effective use of data in decision-making requires "knowledge, skills, and dispositions conducive to systematic gathering, analysis, and interpretation."

While IRAbot affords data retrieval and referential information, analysis is ultimately left to the end-user. IR can offer materials and training that promote data literacy to support the practical analysis of conveyed data. Wolff et al. (2016) define data literacy as "the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data." Data-literate individuals can access, manipulate, and understand the data they are working with (Shields, 2005). Booher-Jennings (2006) cautions that data-driven practices can negatively impact decisions if data are interpreted poorly.

An overabundance of data can cause adverse effects through "information overload" (Orman, 2016). Information overload occurs when information production and availability happen rapidly, possibly making locating required information difficult (Hoq, 2016). Hoq shares common causes for this phenomenon, including too much information, multiple sources of information, information management issues, unnecessary information, and lack of time and context to understand information. Hemp (2014) suggests that technology and cultural improvements can help combat the ailments produced by having too much or irrelevant data. While OIRA publishes self-service data in various forms, data-seekers can be overwhelmed by the amount of information available and options for filtering and searching. Maes (1995) and Edmunds and Morris (2000) identify chatbots as a solution to information overload, citing the ability of AI capabilities to comprehend user requests and route desired data directly to them.

Poor data management practices can present risks in using data to inform decisions. Haug et al. (2011) explain that since organizations often generate and use data, poor quality can affect an organization. Institutions can face financial implications through costs incurred correcting errors or operating with invalid data. Decision-makers may have difficulties achieving strategic goals using inaccurate data (Laranjeiro et al., 2015). Data-seekers may be hesitant to use data they perceive as poor quality. IR and related departments can practice data governance, audits, validation, and publish documentation to instill confidence and trust in their data platforms.

Data Culture. An institution's data culture is guided by the behaviors of collecting, using, and sharing data to inform decisions. Institutional Research can positively impact the health of an institution's data culture by supplying high-quality data and providing insightful analysis to the campus community for decision-making and research purposes. Conversely, poor quality data and inadequate analysis can lead to a negative impact on data culture. Bailey and Alfonso (2005) describe a culture of evidence in which IR "plays a more prominent role and faculty and administrators are more fully engaged with data and research about success of their students, using those data to make decisions." Hamilton et al. (2009) make several recommendations for establishing such a culture, such as making data use for improvement cyclical, establishing a shared vision, providing support for decision-makers, and developing and maintaining data systems.

IR can introduce digital solutions, such as chatbots, into their institution's data economy that support a healthy data culture. Sponsoring self-service options with data pipelines comprised of validated institutional sources and AI capabilities can help IR deliver high-quality data and supporting information to data-seekers without requiring a response from internal personnel. Configuring multiple data sources to flow into a singular gateway breaks down potential silos, improves the democratization of data, and allows for the distribution of consistent data for analysis. Championing data literacy, promoting sound practices, and providing training, documentation, and support ensures that users can interpret and use data from these digital solutions effectively. The improper utilization or implementation of digital solutions can harm an institution's data culture. Poor data quality, data silos, untrained users, unmaintained data solutions, and non-uniform definitions are several hallmarks of an unhealthy data culture.

Chatbots

Chatbots (see Figure 8) are software that interact conversationally with users through voice or text (Smutny & Schreiberova, 2020).

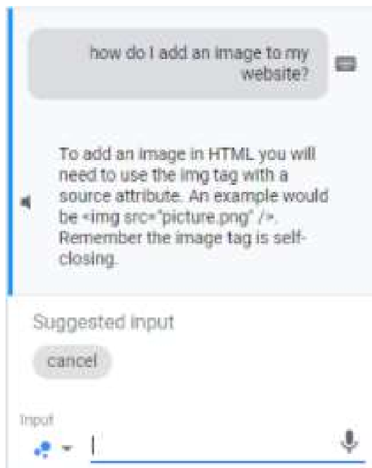

Figure 8. Chatbot message exchange example.

The first chatbot, ELIZA, was programmed in 1966 by Joesph Weizenbaum and used pattern-matching and substitution to converse and act as a therapist for users (Sharma et al., 2017). Today, chatbots can follow predefined scripts or be powered by algorithms and rules designed to perform various tasks beyond simple turn-based conversation. Chatbots can be infused with AI to apply rules, NLP, sentiment analysis, machine learning (ML), and other techniques to mimic natural conversation and perform actions with or for a user. Microsoft Bot Framework and Microsoft Azure work in concert to facilitate these features.

Chatbots have a variety of uses in different industries. Industries employ chatbots to help with shopping, deliver news, book flights, and more (Jain et al., 2018). An institution's data culture can be strengthened by offering a chatbot that features institutional data, actions, and calculations based on common IR requests and practices. Users typically interact with chatbots through text or voice, often in a turn-based format. Many architectures and designs exist for developing chatbots. Molnar and Szuts (2018) describe retrieval-based models and generative models. Retrieval-based models follow a tree structure with user input navigating the branches to arrive at "predetermined dialogs." Generative models do not rely on a predetermined tree structure and instead focus on interaction history and training to generate responses. Many types of chatbots are defined by their purpose and functions, such as informational chatbots, which find and return relevant information using global or local repositories or other resources (Paikari & van der Hoek, 2018).

Waghmare (2019) shares that chatbots feature availability for use beyond regular business hours, possess the capacity to handle multiple conversations, provide a great user experience, metrics on usage, and the potential for work processes to be automated. Using AI, chatbots can determine proper user intent and accurately locate the data and information they are looking for without requiring knowledge in a business intelligence suite or other software. These benefits also complement and can strengthen an institution's data culture by simplifying the process of data gathering for decision-making.

IRAbot

Due to IRAbot having a targeted domain focus and being an informational chatbot, a retrieval-based model was employed. The retrieval-based model allows complete control of dialog and ensures that unapproved outside information or actions do not affect IRAbot's responses. IRAbot includes an interface for users to interact through voice and text, making it a multimodal conversational agent (Jain et al., 2018). A bi-directional approach was taken when designing IRAbot, as it can "both take input and produce output on the communication channel" (Paikari & van der Hoek, 2018), and chatbot-driven dialogue ensures that IRAbot can control the conversation (Følstad et al., 2019). There is no user data retention outside of telemetry log data collected by Microsoft Azure, this relation between the chatbot and user is known as "short-term," and conversations are one-off engagements. The use case diagram for IRAbot is shown in Figure 9 below.

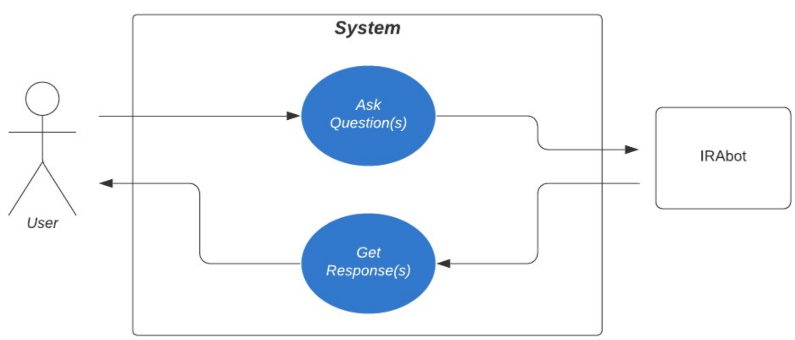

Figure 9. Use Case Diagram.

Combining chatbots and institutional data allows users to gather university information quickly by taking user input, analyzing the request, and returning related data within seconds. Combining multiple data sources can break down data silos and bring data from disparate sources in an organization's data economy onto a centralized platform. Conversational AI offers capabilities beyond a simple data retrieval tool, offering a means for users to ask questions typically reserved for IR analysts. By leveraging the Microsoft Bot Framework, IRAbot was constructed to answer questions surrounding MSU-specific data. Microsoft Azure allows IRAbot to be distributed via a web browser and facilitates communication between linked services. Kowald and Bruns (2020) note that chatbots often face issues with comprehension and contextuality. Chatbots can comprehend and contextualize requests submitted by users and return related information by leveraging the Microsoft Bot Framework and Azure Cognitive Services.

With data consumption, customizability, language understanding through LUIS, advanced search capabilities, and QnA Maker, Microsoft Bot Framework satisfied the conversational AI needs for this project. The data, hosting, and processing services needed were fulfilled using the Microsoft Azure cloud and services. These platforms allow IRAbot to be easily set up and maintained. Actions and flows can be created or modified with Bot Framework Composer or Microsoft Visual Studio and uploaded to Microsoft Azure to replace older versions. New documents and data can be uploaded to Azure Search, an Azure SQL database, or QnA Maker to be parsed by Azure Bot Services. Older documents and data can be purged from the services as needed.

Plan for Future Implementation

Implementation is outside the scope of this capstone. The following is a potential plan that would ensure IRAbot is successfully implemented and that the last three phases of the SDLC are completed. Testing will occur, and a public link to the latest build hosted on Azure will be shared with campus community members. Testing feedback will be collected through a developed email function through IRAbot using SendGrid and Azure Functions. Feedback will be evaluated, and needed changes will be applied to IRAbot. After testing, IRAbot will be redeployed, and the webchat client will be embedded on a public-facing webpage for use. Ongoing maintenance will ensure IRAbot has up-to-date data and any required functionality.

Tentative Timeline:

Completed:

- March 24, 2021: submitted Institutional Review Board (IRB) request for approval to use MSU public-facing data.
- March 25, 2021: requested data from MSU’s OIRA.
- March 29, 2021: received IRB approval for project.
- August 24, 2021: received data from MSU’s OIRA.
- August 25, 2021 – September 30, 2021: parsed received data and scraped MSU webpages for QnA Maker knowledgebases.
- October 1, 2021 – October 31, 2021: researched and developed IRAbot architecture, including data storage, custom development, user interface design, and resource hosting.
- November 1, 2021 – December 31, 2021: built IRAbot.

- January 3, 2022 – January 15, 2022: tested IRAbot with sample queries matching received data requests.
- January 16, 2022 – February 1, 2022: documented build process and results in capstone format.

To be completed:

- Development and Testing: the latest build of IRAbot will be published to Azure, and a URL will be disseminated for testing to members of the campus community. Documentation on functionality and use will be created and shared with users. Feedback generated by participants in the testing group will be evaluated and changes implemented where necessary.
- Deployment: IRAbot will be published with modifications and embedded on public-facing web pages.
- Maintenance: ongoing improvements and data loads.

Intended impact of the capstone

Advantages of IRAbot. The impact of IRAbot is minimal until full implementation. Upon successful implementation, data-seekers will easily find the data they seek on a centralized platform, improving data availability. IRAbot can positively impact institutional data culture by engaging data-seekers and decision-makers who seek and consume information by equipping them with a quick and straightforward way to get the data they need in seconds instead of hours or days. IR analyst and resource usage can be lowered as fewer data requests will need to be filed by users. IRAbot offers extensibility and scalability. With the ability to develop custom functionality and connect to many disparate data sources, IRAbot can support current and future OIRA needs. Azure Bot Analytics gives insight into conversation-level metrics, which allows for exploring user queries and activities that can inform usage trends.

IRAbot implementation would allow OIRA to support data-seekers inside and outside business hours. With a 99.9% monthly uptime Service Level Agreement for Azure Bot Services ("SLA for Azure Bot Services | Microsoft Azure," n.d.), data-seekers can reliably use IRAbot. IRAbot can handle conversations with multiple users simultaneously. Services can be scaled based on demand to save costs when usage is low and introduce more computing resources when usage is elevated.

Disadvantages of IRAbot. IRAbot implementation poses the risk of adverse effects on an organization and users. Waghmare (2019) notes that some disadvantages of chatbots are upfront costs, lack of decision-making aptitude, and level of ability to understand queries and responses. Adamopoulou and Moussiades (2020) share that the need to familiarize users with the chatbot system, lack of personality, and amount of content are also weaknesses of chatbot platforms. To combat these drawbacks, organizational investments can be made. IR offices must promote and support their chatbot platform for full integration into their data ecosystem and culture. Careful planning and implementation can aid in mitigating or overcoming these and other obstacles.

Limitations of the study

Limitations exist that influenced this project. The limited audience and scope of this capstone hamper generalizability. Without modification, IRAbot cannot be used at other institutions. Currently, IRAbot only houses a subset of MSU public data and may not feature the information a user is seeking. The language understanding capabilities of IRAbot are trained at a low volume and may not be able to interpret all user queries appropriately. Time constraints reduced the breadth of the project. Safety aspects and the potential for AI to deviate from intended purposes due to malicious feeds or actions (Zanetti et al., 2019) influenced the design and architecture of IRAbot. Trust and platform adoption will need to be considered to establish that IRAbot can produce accurate data and promote availability. Documentation at this point is not mature and will need to be improved. Data security and legal requirements will need to be addressed moving forward as data use and storage are expanded with IRAbot. As technology evolves, techniques and tools used to build IRAbot are at risk of depreciation by their respective software vendors or maintainers.

Reflections

With chatbot utilization becoming more common across higher education, I believe that IRAbot showcases that IR and IT offices can offer an improved user experience, increase data availability and accessibility, save resources, and automate certain functions through conversational AI. IRAbot demonstrates successful data retrieval in seconds, whereas some IR office practices can take hours or days for returns, such as traditional data request methods. While IRAbot was designed to decrease resource usage and promote automation, it is meant to supplement humans and not replace them. De Cremer and Kasparov (2021) acknowledge that current AI offerings do not have the capacity to fully replicate or supplant human knowledge and functions.

Learning new tools and technologies can prove to be time-consuming. While I enjoy working with unfamiliar tools, I feel my proposed timeline and outside influences stalled opportunities to take full advantage of the Microsoft Bot Framework platform offerings. I intend to revisit the platform's functionality and revise IRAbot as appropriate. Developing and implementing technology tools requires sound project management to avoid having unforeseen circumstances hinder progress.

IRAbot design has been informed by reviewing published research, data requests, and experience. Moving forward, I believe more input is needed from users to eliminate potential barriers to data and refine interactions and usability. Too often, technology projects fail due to a lack of focus on essential components of a project, such as communication or engagement with end-users. Keeping the audience and project goals in mind is paramount when developing or implementing technology.

While I initially composed this project as an avenue to introduce a budding technology into OIRA, research and discussions with my doctoral committee helped broaden the intended impact. After reflecting on my approach, I decided to appropriate more time to work with my audience beyond this capstone. Without proper problem framing and exploration, I would have proposed implementing a "technology for technology's sake" (Bensaou & Earl, 1998). I aim to keep these considerations constant when exploring or crafting technology solutions in the future.

Capstone Project

The capstone project is represented in this section through screenshots of user interaction, sample queries, and code snippets. A video demonstration is available at https://youtu.be/WtW8zyzpSeM.

- Figure 10 showcases IRAbot greeting a user, receiving input from the user, and generating a response.

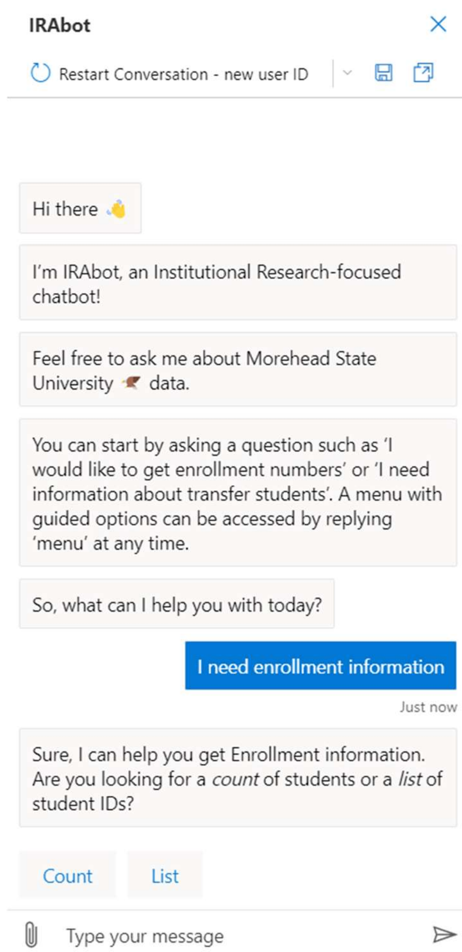

Figure 10. IRAbot greeting and response.

- Figure 11 displays IRAbot's response to an unknown request.

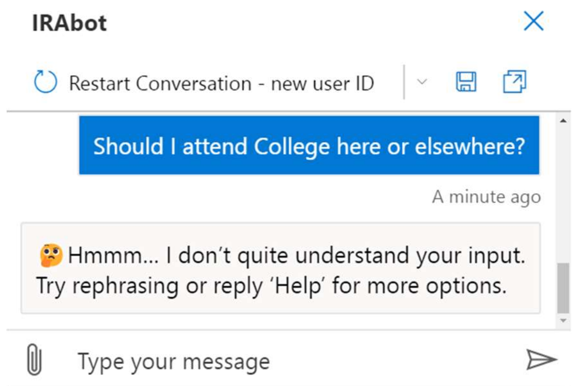

Figure 11. Response to unknown phrasing.

- Figure 12 demonstrates IRAbot returning results based on a user request.

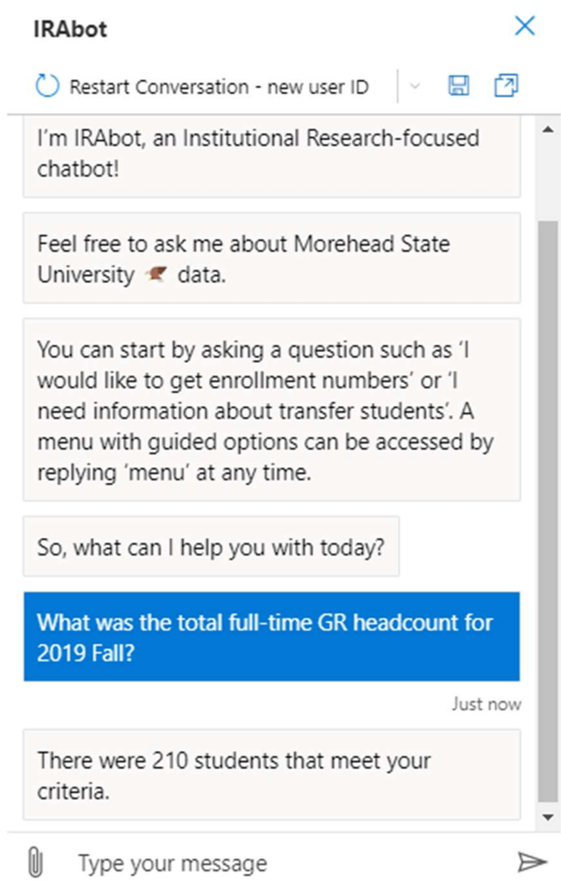

Figure 12. Result response based on a user request.

- Figure 13 features IRAbot returning Transfer Student results based on a guided query.

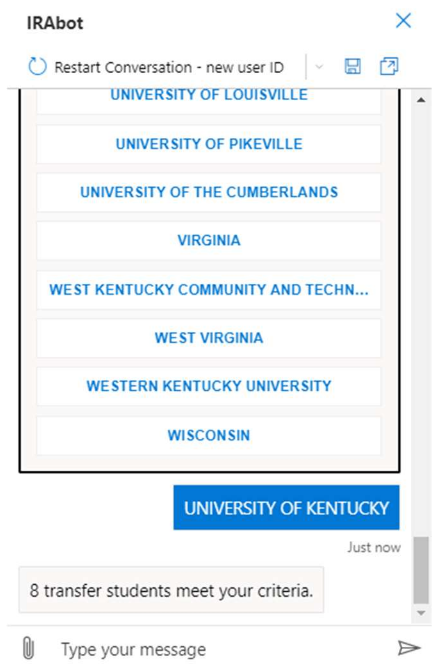

Figure 13. Transfer student result based on a guided query.

- Figure 14 exhibits IRAbot returning Graduating Student results based on a guided query.

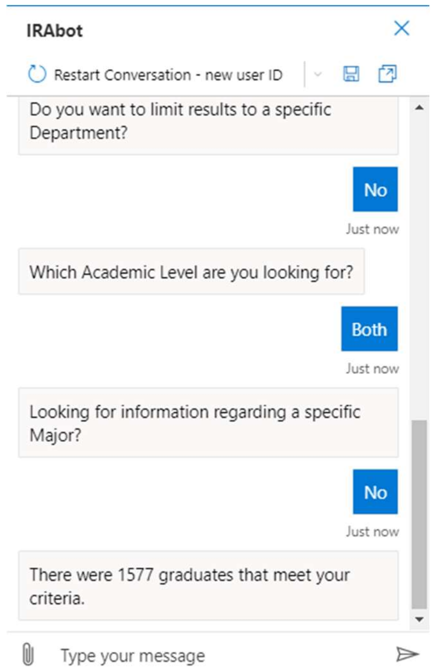

Figure 14. Graduating student result based on a guided query.

- Figure 15 illustrates IRAbot returning a list of student IDs.

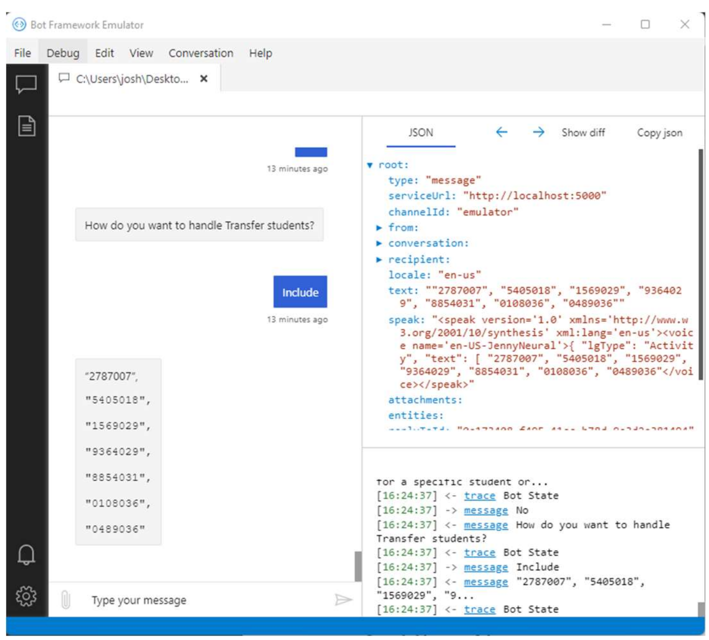

Figure 15. IRAbot returning a list of student IDs.

- Figure 16 showcases IRAbot returning results to a user query with information parsed from the MSU About webpage.

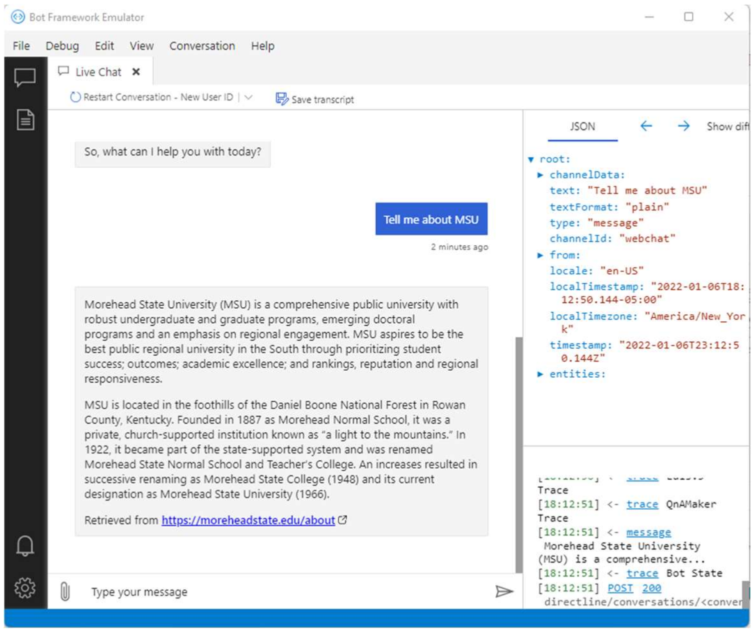

Figure 16. General MSU information scraped from the MSU website by IRAbot.

- Figure 17 displays IRAbot returning results to a user query with information parsed from the MSU About webpage.

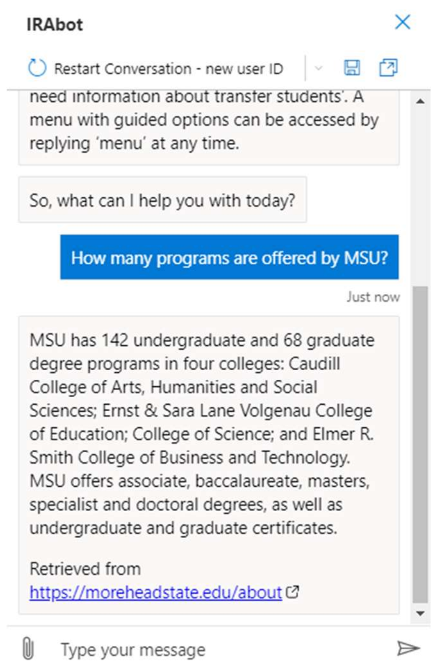

Figure 17. General program information scraped from the MSU website by IRAbot.

- Figure 18 demonstrates IRAbot's 'Explain' functionality.

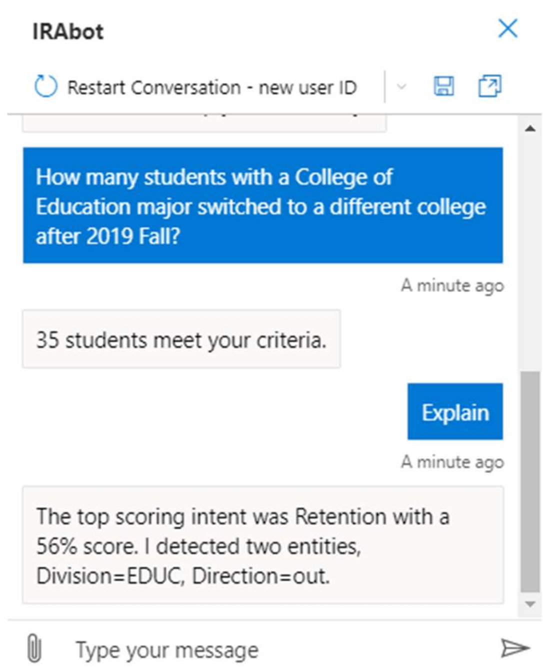

Figure 18. IRAbot detailing intent and entity information.

- Figure 19 shows the dialog flow for ‘Help’ requests.

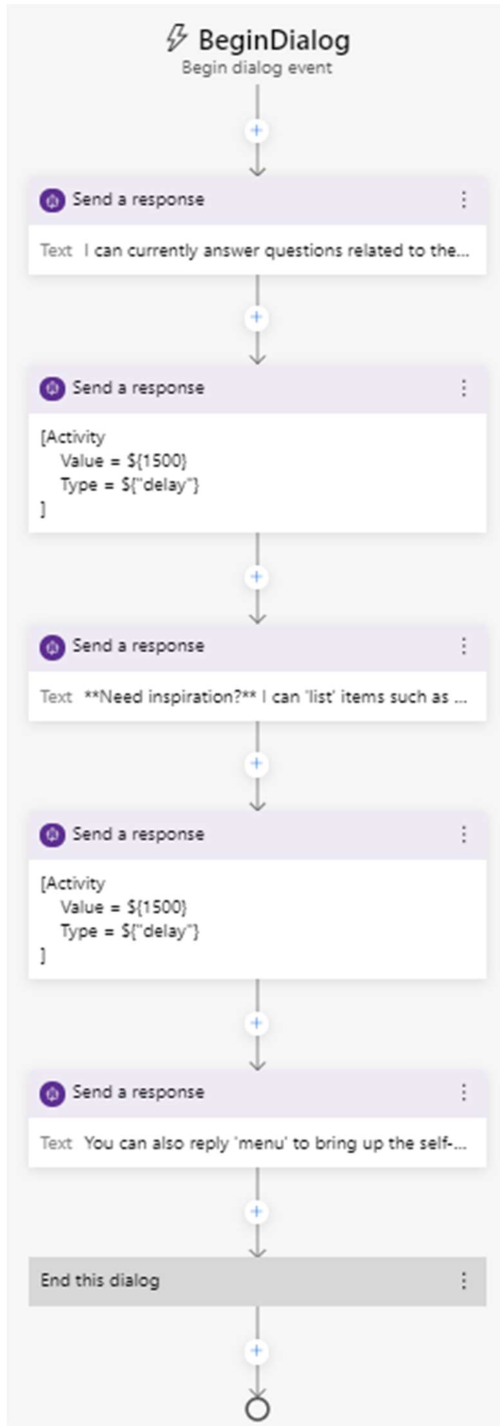

Figure 19. Dialog flow for help requests.

- Figure 20 features dialog flow for guided Transfer Student queries.

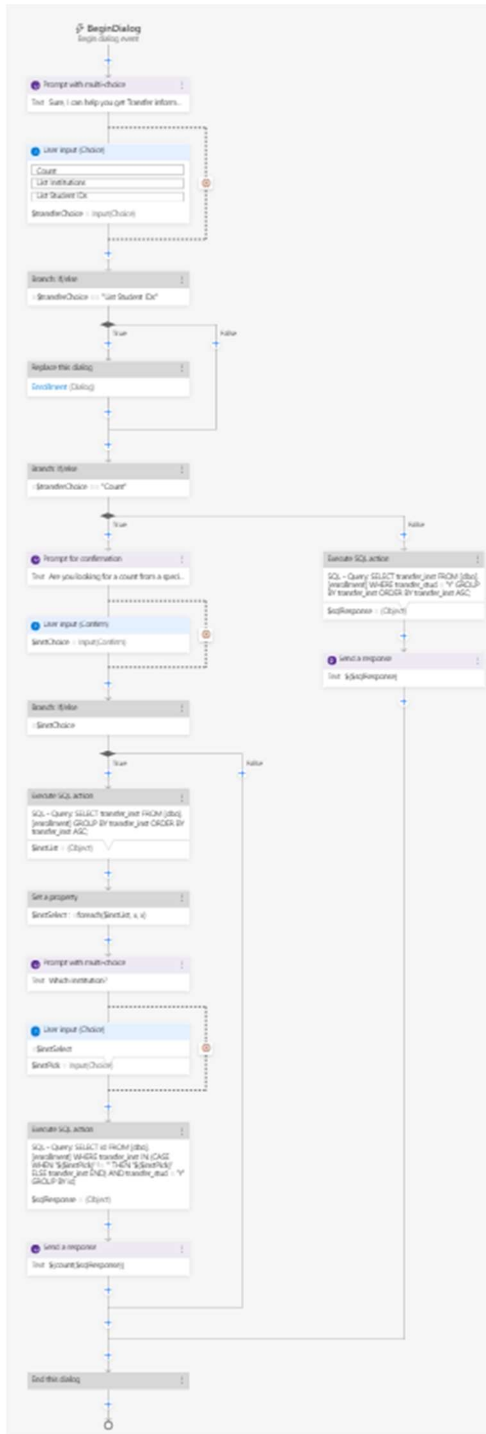

Figure 20. Dialog flow for transfer student requests.

- Figure 21 exhibits the SQL statement to return guided Enrollment results.

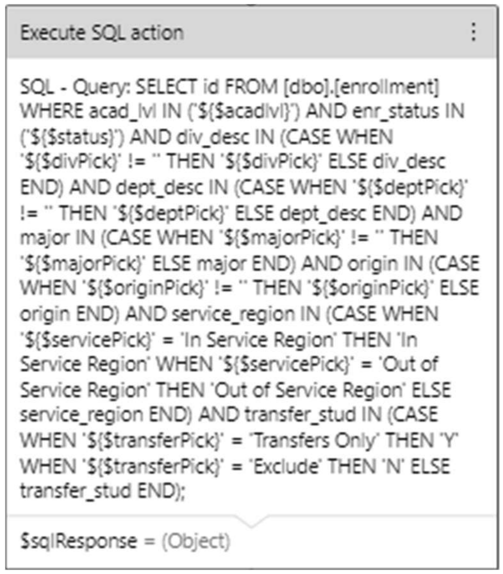

Figure 21. SQL statement for guided enrollment results.

- Figure 22 illustrates Microsoft Azure portal view of the IRAbot Azure SQL database with an executed query.

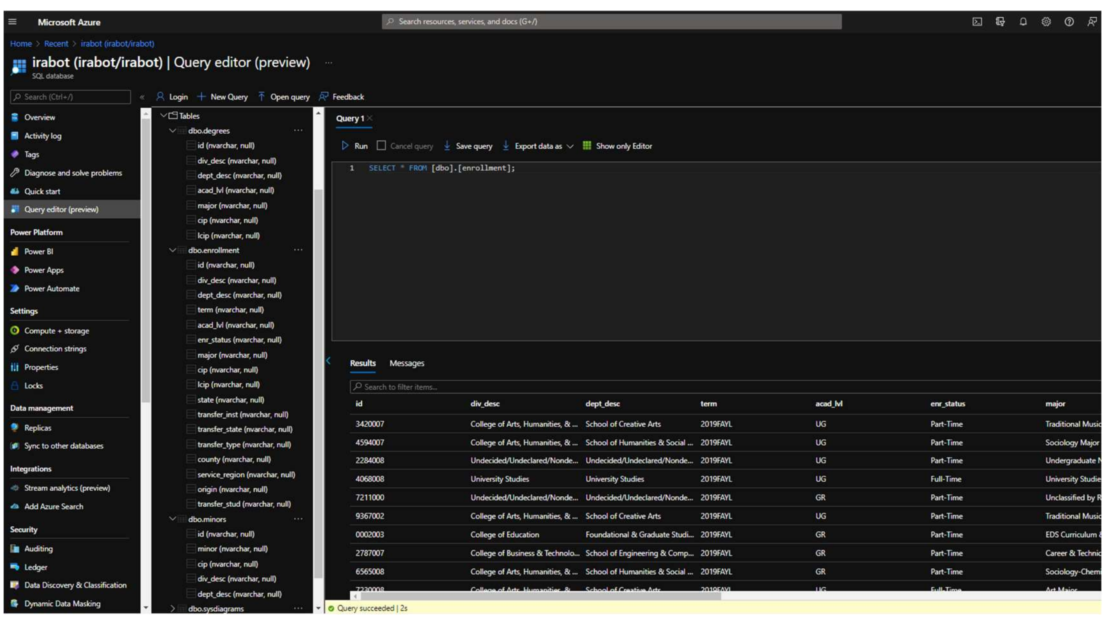

Figure 22. IRAbot Azure SQL query view.

REFERENCES

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, *2*, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Agrawal, P., Menon, T., Kam, A., Naim, M., Chouragade, C., Singh, G., … Rajput, N. (2020). QnAMaker: Data to Bot in 2 Minutes. *Companion Proceedings of the Web Conference 2020*. https://doi.org/10.1145/3366424.3383525

Albright, S. C., Winston, W. L., Zappe, C. J., & Broadie, M. N. (2011). *Data analysis and decision making* (Vol. 577). South-Western/Cengage Learning.

Bailey, T. R., and Alfonso, M. (2005). Paths to Persistence: An Analysis of Research on Program Effectiveness at Community Colleges. *New Agenda Series*, 6(1). Indianapolis, IN: Lumina Foundation for Education.

Barrett, M., Branson, L., Carter, S., DeLeon, F., Ellis, J., Gundlach, C., & Lee, D. (2019). Using artificial intelligence to enhance educational opportunities and student services in higher education. *Inquiry: The Journal of the Virginia Community Colleges*, *22*(1).

Booher-Jennings, J. (2006). Rationing education in an era of Accountability. *Phi Delta Kappan*, *87*(10), 756–761. https://doi.org/10.1177/003172170608701010

Bensaou, M., & Earl, M.E. (1998). The right mind-set for managing information technology. *Harvard business review, 76 5*, 118-28, 188.

Brandtzaeg, P. B., & Følstad, A. (2017). Why People Use Chatbots. *Internet Science*, 377–392. https://doi.org/10.1007/978-3-319-70284-1_30

Chaves, A. P., & Gerosa, M. A. (2020). How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human–Computer Interaction*, *37*(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438

Chowdhury, G., & Lynch, M. F. (1991). Natural language processing of the texts of chemical patent abstracts. In *Intelligent Text and Image Handling-Volume 2* (pp. 740-753).

Coburn, C. E., & Turner, E. O. (2012). The practice of Data use: An introduction. *American Journal of Education*, *118*(2), 99–111. https://doi.org/10.1086/663272

Dahiya, M. (2017). A tool of conversation: Chatbot. *International Journal of Computer Sciences and Engineering*, *5*(5), 158–161.

Data Analysis. (n.d.). Retrieved April 8, 2022, from https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/datopic.html

Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: A review of the literature. *International Journal of Information Management*, *20*(1), 17–28. https://doi.org/10.1016/s0268-4012(99)00051-1

De Cremer, D., & Kasparov, G. (2021). AI should augment human intelligence, not replace it. *Harvard Business Review*.

Díaz, A., Rowshankish, K., & Saleh, T. (2018). Why data culture matters. McKinsey Quarterly, 3(1), 36-53.

Fontichiaro, K., & Oehrli, J. A. (2016). Why Data Literacy Matters. *Knowledge Quest*, *44*(5), 21–27.

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2019). Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. *Internet Science*, 145–156. https://doi.org/10.1007/978-3-030-17705-8_13

Gill, B., Borden, B. C., & Hallgren, K. (2014). A conceptual framework for data-driven decision making. *Final Report of Research conducted by Mathematica Policy Research, Princeton, submitted to Bill & Melinda Gates Foundation, Seattle, WA*.

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society*, *15*(3), 42–57.

Groff, J., & Weinberg, P. (2010). *Sql, the complete reference*. McGraw-Hill.

Gupta, A., Hathwar, D., & Vijayakumar, A. (2020). Introduction to AI Chatbots. *International Journal of Engineering Research And*, *V9*(07). https://doi.org/10.17577/ijertv9is070143

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J. A., & Wayman, J. C. (2009). *Using student achievement data to support instructional decision making*. U.S. Dept. of Education, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data Availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *cognition*. *Royal Society Open Science*, *5*(8), 180448. https://doi.org/10.1098/rsos.180448

Haug, A., Zachariassen, F., & Liempd, D. van. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, *4*(2). https://doi.org/10.3926/jiem.2011.v4n2.p168-193

Hemp, P. (2014, August 1). *Death by information overload*. Harvard Business Review. Retrieved April 10, 2022, from https://hbr.org/2009/09/death-by-information-overload

Hoq, K. M. (2016). Information overload: Causes, consequences and remedies - A study. *Philosophy and Progress*, 49–68. https://doi.org/10.3329/pp.v55i1-2.26390

Huda, M., Anshari, M., Shahrill, M., & Almunawar, M. (2016). Innovative Teaching in Higher Education: The Big Data Approach. *Turkish Online Journal of Educational Technology*. https://doi.org/10.13140/RG.2.2.10543.53924

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. *Proceedings of the 2018 Designing Interactive Systems Conference*. https://doi.org/10.1145/3196709.3196735

Kennedy, H., Oman, S., Taylor, M., Bates, J., & Steedman, R. (2021). Public Understanding and Perceptions of Data Practices: A Review of Existing Research. University of Sheffield. http://livingwithdata.org/current-research/publications/

Kowald, C., & Bruns, B. (2020). Chatbot Kim: A digital tutor on AI. How advanced dialog design creates better conversational learning experiences. *International Journal of Advanced Corporate Learning (IJAC)*, *13*(3), 26. https://doi.org/10.3991/ijac.v13i3.17017

Laranjeiro, N., Soydemir, S. N., & Bernardino, J. (2015). A survey on data quality: Classifying Poor Data. *2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC)*. https://doi.org/10.1109/prdc.2015.41

Letzring, T. D. (2017). What Did You Know and When Did You Know It? Data's Potential for Institutional Legal Issues. *New Directions for Institutional Research*, *2016*(172), 61–71. https://doi.org/10.1002/ir.20204

Leu, K. (2020). Data for Students: The Potential of Data and Analytics for Student Success. https://doi.org/10.3768/rtipress.2020.rb.0023.2003

Louch, M., & Pry, M. (2020). Ethics and Data Manipulation. *Information Systems Education Journal*, *18*(2), 4–13.

Maes, P. (1995). Agents that reduce work and information overload. *Readings in Human–Computer Interaction*, 811–821. https://doi.org/10.1016/b978-0-08-051574-8.50084-4

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71–85. https://doi.org/10.1080/00461520.2012.667064

Merriam-Webster. (n.d.). *Data Definition & meaning*. Merriam-Webster. Retrieved April 8, 2022, from https://www.merriam-webster.com/dictionary/data

Merriam-Webster. (n.d.). *Database definition & meaning*. Merriam-Webster. Retrieved April 8, 2022, from https://www.merriam-webster.com/dictionary/database

Merriam-Webster. (n.d.). *Frequency distribution definition & meaning*. Merriam-Webster. Retrieved April 8, 2022, from https://www.merriam-webster.com/dictionary/frequency%20distribution

McCarthy, J. (2007). What is artificial intelligence?.

Microsoft Bot Framework. (n.d.). https://dev.botframework.com/.

Molnar, G., & Szuts, Z. (2018). The Role of Chatbots in Formal Education. *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*. https://doi.org/10.1109/sisy.2018.8524609

Moniz, A., Gordon, M., Bergum, I., Chang, M., & Grant, G. (2021). *Beginning azure cognitive services: Data-driven decision making through Artificial Intelligence*. Apress.

Morehead State University :: About Morehead State University. (n.d.). https://www.moreheadstate.edu/About-MSU.

Morehead State University :: Morehead State University Administration. (n.d.). https://www.moreheadstate.edu/ira.

Morehead State University :: Morehead State University. (n.d.). https://www.moreheadstate.edu/Administration/PPE/Institutional-Research/Institutional-Data.

Murray, J. (2013). Critical issues facing school leaders concerning data-informed decision-making. *School Leadership & Management*, *33*(2), 169–177. https://doi.org/10.1080/13632434.2013.773882

Naqa, I., Li, R., & Murphy, M. (2015). *Machine learning in Radiation oncology: Theory and applications*. Springer.

Nilsson, N. (2009). *The quest for artificial intelligence*. Cambridge University Press.

Orman, L. V. (2016). *Information overload paradox: Drowning in information, starving for knowledge*.

Paikari, E., & van der Hoek, A. (2018). A framework for understanding chatbots and their future. *Proceedings of the 11th International Workshop on Cooperative and Human Aspects of Software Engineering*. https://doi.org/10.1145/3195836.3195859

Picciano, A. (2019). Artificial Intelligence and the Academy's Loss of Purpose. *Online Learning*, *23*(3). https://doi.org/10.24059/olj.v23i3.2023

Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017). Chatbot for university related FAQs. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. https://doi.org/10.1109/icacci.2017.8126057

Reeves, P. L., & Burt, W. L. (2006). Challenges in data-based decision-making: Voices from principals. *Educational Horizons*, *85*(1), 65-71.

Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, *26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Saadia, D. (2021). Integration of Cloud Computing, Big Data, Artificial Intelligence, and Internet of Things. *International Journal of Web-Based Learning and Teaching Technologies*, *16*(1), 10–17. https://doi.org/10.4018/ijwltt.2021010102

Saupe, J. L. (1990). *The functions of Institutional Research*. Association for Institutional Research.

Sharma, V., Goyal, M., & Malik, D. (2017). An Intelligent Behaviour Shown by Chatbot System. *International Journal of New Technology and Research*, *3*(4), 52–54.

Shields, M. (2005). Information Literacy, statistical literacy, Data Literacy. *IASSIST Quarterly*, *28*(2), 6. https://doi.org/10.29173/iq790

Singh, A., Ramasubramanian, K., & Shivam, S. (2019). *Building an enterprise chatbot: Work with Protected Enterprise Data Using Open Source Frameworks*. Apress.

SLA for Azure Bot Services | Microsoft Azure. (n.d.). https://azure.microsoft.com/en-us/support/legal/sla/bot-services/v1_0/

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, 103862. https://doi.org/10.1016/j.compedu.2020.103862

Tamayo, P. A., Herrero, A., Martín, J., Navarro, C., & Tránchez, J. M. (2020). Design of a chatbot as a distance learning assistant. *Open Praxis*, *12*(1), 145. https://doi.org/10.5944/openpraxis.12.1.1063

Turk, J. M., & Taylor, M. (2019). Institutional Research in Support of Student Success at Our Nation's Most Diverse and Inclusive Institutions. *New Directions for Institutional Research*, *2019*(184), 75–90. https://doi.org/10.1002/ir.20323

Waghmare, C. D. (2019). *Introducing azure bot service: Building bots for business*. Apress.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

What is FERPA? What is FERPA? | Protecting Student Privacy. (n.d.). Retrieved April 11, 2022, from https://studentprivacy.ed.gov/faq/what-ferpa

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3275

Zanetti, M., Iseppi, G., & Cassese, F. P. (2019). A “psychopathic” Artificial Intelligence: the possible risks of a deviating AI in Education. *Research on Education and Media*, *11*(1), 93–99. https://doi.org/10.2478/rem-2019-0013

VITA

JOSHUA C. FRISBY

EDUCATION

July, 2013	Bachelor of Business Administration Morehead State University Morehead, Kentucky
May, 2016	Master of Science Morehead State University Morehead, Kentucky
Pending	Doctor of Education Morehead State University Morehead, Kentucky

PROFESSIONAL EXPERIENCES

2022-Present	Enterprise Developer University of Kentucky Lexington, Kentucky
2021-Present	Adjunct Instructor, Computer and Information Sciences University of the Cumberlands Williamsburg, Kentucky
2021-2022	Sr. Systems Analyst Transylvania University Lexington, Kentucky
2017-2021	Programmer Analyst, Institutional Research Morehead State University Morehead, Kentucky
2017	Senior IT Auditor Pikeville Medical Center Pikeville, Kentucky
2016	Instructor, Computer Information Systems Morehead State University Morehead, Kentucky

2014-2016	Technology Business Analyst II (ERP) Morehead State University Morehead, Kentucky
2013-2014	IT Manager The Gallaher Group Ashland, Kentucky
2011-2013	IT Specialist N.E. Kentucky Regional Health Information Organization West Liberty, Kentucky