Chatbot Assessment: Best Practices for Artificial Intelligence in the Library

Michelle Ehrenpreis and John DeLooper

abstract: In November 2019, the Leonard Lief Library implemented Ivy.ai, a proprietary chatbot on its website. This implementation was the first academic library installation of a vendor-supplied chatbot to be discussed in the professional literature. This chatbot functioned as a new tool that assisted users seeking information from the library website. User questions provided insight to the authors about the kinds of topics students searched for via the library website. In April 2023, the chatbot's vendor began using OpenAI's ChatGPT Application Programming Interface (API) to improve the chatbot's functionality. This change, from a rules-based chatbot system to a transformer model, enhanced the chatbot's ability to provide answers to patrons. To better understand this major change, the authors assessed the chatbot's usage during the Spring 2023 semester. This assessment revealed the kinds of questions the chatbot struggled to answer, and possible reasons why. The assessment's findings demonstrated how chatbots can successfully function as a enhancement to the library website. The article also presents best practices for libraries looking to implement or experiment with chatbots and contributes to the ongoing discussion of artificial intelligence in libraries.

Introduction

Chatbots

A chatbot is a computer program designed to have a conversation with a human being, usually over the internet. Chatbots generate text responses, often via the artificial intelligence (AI) technique of training on large amounts of information, called a knowledge base. Chatbots that use this technique are called large language model (LLM) chatbots.[1] Chatbots can be found on many types of websites including e-commerce, education, and social media platforms. In November, 2022, OpenAI released ChatGPT, a free-to-use AI system, which was used for "engaging in conversations,

portal: Libraries and the Academy, Vol. 25, No. 4 (2025), pp. 669–701.

gaining insights, automating tasks, all in one place."[2] What made ChatGPT particularly special compared to previous iterations of chatbots, was that "the quality of the outputs [of previous chatbots] was much lower than that produced by an average human. The new model is much better, often startlingly so. Put simply: This is a very big deal."[3] Within days of its release to the public, eager testers experimented with a bevy of tasks for ChatGPT such as writing Python scripts or creating new online content.[4] ChatGPT's speed in replying to user queries and synthesizing information into compact, readable text appealed to users who could apply it to suit their needs. This represented a large step forward from using a search engine like Google and having to scroll through links to obtain the information.

Library Setting the Ivy Chatbot

The authors conducting this research comprise the Electronic Resources and Web Services-Online Learning Unit, and one of the authors manages the Leonard Lief library's website and software. Both authors are administrators of the Ivy Chatbot and manage the back end and front end of the system. The Leonard Lief Library is Lehman College's library and is part of the City University of New York (CUNY) consortium, which consists of twenty-five colleges across the city of New York. Lehman College is a four-year college, which has eighty undergraduate and sixty-six master's degree programs in nursing, psychology, business administration, health services administration, and sociology. The library serves approximately 14,000 full-time students, many of whom take online classes or are enrolled in online or hybrid master's or doctoral programs.[5] While six other colleges in CUNY also use Ivy as their college's chatbot, no other library uses the Ivy chatbot at the time of writing this article.

In 2019, The Leonard Lief Library at Lehman College implemented Ivy.ai's chatbot alongside other college departments including the registrar, financial aid, bursar, IT, and the college's learning management system, Blackboard. Ivy.ai (also called Ivy) is a software company that licenses its chatbot to higher education, healthcare, state and local government industries.[6] Implementation of the chatbot software was managed under the direction of the IT department.[7] The library's chatbot instance was shared with the college's Blackboard site, which allowed the authors and the Blackboard administrator to collaborate on issues that affected both units.

To prepare for the implementation, the library supplied IT with sample questions that the chatbot would likely be asked, such as, "What are the library hours?" These questions were developed based on questions the library received frequently at the reference desk. Five library webpages were supplied to Ivy for the chatbot to crawl and build its "brain." This "brain" was a knowledge base of information from which the chatbot would glean answers. The implementation of Ivy's software resulted in several website updates and created a framework for the chatbot to function as an online tool for the library to answer users' questions, alongside the library's online 24/7 chat.[8] The authors also enabled Ivy's built-in ticketing system, that enabled users to fill out a form with their contact information and seek help with questions that were not answered to their satisfaction. This form was routed directly to the authors via email. At the time of publication, the chatbot remains a staple on the library website, situated on the homepage and a dedicated chatbot page, available for 24/7 use.

Pre-Ivy Quantum Chatbot Impressions in 2020

After the chatbot's implementation in Fall 2019, the authors reviewed chatbot tickets from the Spring 2020 semester. This review showed three primary categories of questions: textbooks, off-campus access to library databases, and research-related questions.[9] In response to these findings, the library created a "remote resources" research guide that detailed free access to textbooks and database access to eBook platforms available from the library. The guide also provided information about off-campus access to the library databases, a troubleshooting ticket form, links to the library's Ask-a-Librarian chat service, and the IT department.[10] The presence of research questions surprised the authors, who concluded at the time that as long as the volume of research questions received was manageable and went primarily to the library's online chat service instead of the chatbot, they could continue to be answered via the ticketing system.[11]

The implementation demonstrated the chatbot's difficulty in crawling the library website, so Author B redesigned the website layout in Fall 2020 to allow for better crawling by Ivy.[12]

IvyQuantum Chatbot

In April 2023, Ivy launched their IvyQuantum chatbot, which was powered by Open AI's ChatGPT 3.5 Application Programming Interface (API). An API is a process for allowing different software programs to communicate and exchange data. In this case, the addition of the ChatGPT API allowed the IvyQuantum chatbot to gain access to ChatGPT's software and pass data back and forth with it. Ivy's chatbot also incorporated a knowledge base made up of data found on the library's website. As users asked the chatbot questions, IvyQuantum was able to pass data to ChatGPT and utilize its own data from the library website, which it would use for verification as part of its generation process.[13]

Data Retrieval Process

The process of combining data from a chatbot with data found in outside sources is known as Retrieval Augmented Generation (RAG).[14] Companies like Ivy use RAG to supplement the data used by a chatbot (in this case, the data found in ChatGPT's model and its training data) with data that is found outside of a chatbot's training data (in this case, the information published on the library website) as an additional knowledge base. This is especially important since ChatGPT's data is generally a few years old and may not include many websites. By adding in a local knowledge base like the library website, a chatbot like Ivy Quantum could provide "facts" to supplement the responses a large language model like ChatGPT generates. Furthermore, because RAG checks against an additional knowledge base of current data, this model can better avoid hallucinations and thus provide more accurate information to end users.[15]

By adding in a local knowledge base like the library website, a chatbot like Ivy Quantum could provide "facts" to supplement the responses a large language model like ChatGPT generates.

Combining Knowledge Bases

Even without the library's local knowledge base being used by the chatbot, OpenAI's ChatGPT software utilized a process for answering questions known as a transformer model, which responded to queries using a method called "attention." The implementation of attention has been demonstrated to improve a chatbot's ability to retrieve information.[16] By utilizing ChatGPT's technology stack, the IvyQuantum chatbot was able to provide generative answers for many questions it received. Generative answers are essentially "new" content that is created without human intervention.[17] This differed from the previous Ivy model, which was a rules-based chatbot and utilized pre-defined rules and keywords to respond to questions from users with pre-defined answers (sometimes called a dialogue tree).[18]

However, even though Ivy had added the ChatGPT transformer technology to its product, IvyQuantum also continued to utilize rules-based responses to questions when the librarians had already supplied information, effectively making IvyQuantum a hybrid of generative and rules-based models.

Additional Upgrades

As part of the Retrieval Augmented Generation process, IvyQuantum also introduced the ability for its chatbot to crawl the entire library website daily, without additional upcharge. This contrasted with the college's original agreement with Ivy, which allowed for only five webpages to be crawled. The expanded amount of information crawled by the chatbot allowed it to utilize a total of 162,495 pages as of publication. This included not only the library website, but also its detailed research guides, list of databases, FAQs, and more.

With the integration of a transformer model and the addition of an expanded knowledge base to allow RAG, IvyQuantum could respond more dynamically to user questions, rather than pulling from a set of pre-defined responses, to generate appropriate answers. The authors initially thought that the chatbot would also engage in a high degree of learning based on interactions with patrons and add these interactions to its knowledge base (and possibly to ChatGPT). However, Ivy's agreement with Lehman College did not allow this. Nevertheless, the generative transformer technology was able to utilize the larger dataset provided through access to ChatGPT's API and the expanded set of webpages crawled.

Thanks to the expanded dataset and the ChatGPT technology, the chatbot could provide a greater degree of back-and-forth interaction, as it provided answers to satisfy each patron when they asked questions and introduced clarifying questions. For the Leonard Lief library, this manifested as an upgrade to the chatbot's ability to answer user questions in detail and made the user experience feel very similar to that of using ChatGPT, which many library users were experimenting with at the time.

Finally, with the new interface, administrative users were also granted greater ability to customize rules-based responses to commonly asked questions as they occurred. This represented a significant upgrade for ease of maintenance, as the process no longer required sending support tickets to Ivy for each interaction that the chatbot could not answer.

Library's Chatbot Assessment

Given the major upgrades discussed with IvyQuantum, the authors decided in Fall 2023 that a formal assessment of the chatbot was necessary to evaluate the efficacy of the chatbot in meeting the information needs of the Lehman college community. Specifically, the authors were looking to see what kinds of questions the chatbot struggled to answer, in subject areas such as textbook requests, library hours, and more. Based on the assessment findings, the authors hoped to make improvements to the library website and chatbot platform so that these questions could be answered more reliably in the future.

To conduct this assessment, the authors proposed a case study in which they would review a random sample of questions received during the Spring 2023 semester. This random sample consisted of 101 questions out of the 816 questions received during that period. The authors also proposed studying a purposive sample of all thirty-nine tickets received by the library during Spring 2023. The tickets were questions that the chatbot could not answer and therefore, prompted users to fill out a form with their name, email address, and question. The authors received these forms via email in order to provide direct responses to the users. A series of variables was determined based on built-in Ivy-supplied metrics for success and failure and author-constructed variables. These variables formed the basis of a rubric that was used to score transcripts and tickets for accuracy and completeness. The authors then calculated how many questions from the random sample and ticket sample were answered correctly or incorrectly, and completely or incompletely, based on the rubric. The authors created this rubric using several sources, including the Reference and User Services Association (RUSA) guidelines, which provided methods to successfully conduct reference interviews, as a model. Studies that evaluated online chat reference were used as a reference for developing user behavior variables to measure patron satisfaction as one indication of answer accuracy. The authors then conducted a content analysis of the transcripts and tickets to identify topics users queried the chatbot about. Each query was then coded into descriptive terms. The total number of questions that stumped the chatbot and the topics of these questions were then collated to determine exactly where the chatbot was struggling. These challenges could then be addressed after the assessment.

Research Goals

Through this assessment, the authors hoped to achieve several research goals. These included pinpointing questions where the chatbot failed to answer patrons correctly and finding specific ways in which the chatbot knowledge base could be amended so that these questions would be reliably answered in future interactions. The authors also sought to identify improvements that could be made to the library's website, in terms of content or structure, that would enable the chatbot to more easily find relevant information when answering their questions. By achieving these goals, the authors hoped to formulate a long-term plan to maintain and improve the chatbot. This assessment

This assessment was the first of its kind to measure a library chatbot which used a licensed ChatGPT API, rather than assessing ChatGPT directly based on a series of tasks.

was the first of its kind to measure a library chatbot which used a licensed ChatGPT API, rather than assessing ChatGPT directly based on a series of tasks. The rubric the authors created was also unique as a formal means of assessing a chatbot, taking into consideration user satisfaction, the platform's metrics for success, and librarian review of the chatbot's answers to patrons' questions. In this article, the authors also highlight some advantages of this model as an alternative to using the free version of ChatGPT, which became evident after reviewing the assessment's findings and attempting to implement them. The authors have shared their learned experience and presented best practices for other libraries that may wish to experiment with chatbots. This is particularly important because the use of Artificial Intelligence tools in libraries appears to be increasing, and many librarians have worked to develop applications and are interested in their long-term effects.

Literature Review

Library Chatbot Assessments

Prior to ChatGPT, several libraries created their own chatbots and assessed them, focusing primarily on popular questions, categories of topics, and setup. None of these studies included any form of rubric for scoring their chatbots' interactions with users for accuracy and completeness, an important metric for improving the chatbot's performance.

University of California, Irvine (UCI) assessed their chatbot, ANTswers, from its inception in March, 2014 to April, 2018 to determine how patrons asked questions to the chatbot and what types of resources and services they asked about. UCI's researchers sorted questions into broad categories such as About (the library), About (UCI), Find, Services, and Subject, and then into an appropriate narrow category. The study's findings confirmed what UCI suspected—most library patrons asked directional questions and other simple questions about library services, locations, and policies. By and large, patrons using the chatbot did not ask in-depth research questions. Based on the number of questions ANTswers received about library hours, UCI placed hours information in a prominent position on the main page of their website. This was their only website enhancement based on their chatbot implementation.

San Jose State University (SJSU) Library assessed their Kingbot chatbot after eighteen months of use, categorizing transcripts of chat sessions into topics such as building information, live reference hours, finding peer-reviewed articles, research help, and requests to speak with a librarian.[19] SJSU's transcript review showed several important limitations in their Dialogflow chatbot software, namely that user clicks, heat maps, and scroll maps could not be recorded. SJSU's authors noted that analyzing these data would have been important to understanding how users proceeded with the information provided by the chatbot. In other words, these data would be required to determine whether users trusted what the chatbot was telling them or not. To address this issue, SJSU inserted suggestion chips, "or predefined buttons the users may select through an automated prompt. Suggestion chips might display a question such as: 'Did I help you find what you were looking for?' followed by a 'Yes' or 'No' response selection."[20] These suggestion chips prompted the user for feedback on the interaction. The SJSU authors hoped that this feedback would lead to suggestions for improving conversation flows and responses.

Library Assessments of ChatGPT

After ChatGPT debuted, many libraries were eager to test out its capabilities, particularly in anticipation of students using it for research papers. The literature discussed experimentation with the tool and offered limited evaluations based on these experiences but did not provide comprehensive assessments based on users' actual questions posed to ChatGPT. The research questions were often based on popular chat reference questions.

Katie Lai asked ChatGPT 58 questions that had been received by a music library's reference email. ChatGPT's answers were analyzed using a three-point rubric for accuracy, completeness, and further assistance.[21] Lai's study highlighted ChatGPT's strengths, which included politeness, multilingual capabilities, and the tool's ability to build on provided information, such as the user's affiliation with the school, to tailor its answers. Lai's article also showcased ChatGPT's limitations, like failure to detect nuance, trouble referring to other sources, and an inability to search outside its training data. The study was limited to questions relating to a specialized library with only one author coding the data.

Yrjo Lappalainen and Nikesh Narayanan of Zayed University Library (United Arab Emirates) created a custom chatbot, Aisha, using Python and the ChatGPT API to support student and faculty simple reference needs outside normal business hours. Without access to the library website or plugins to outside websites, Aisha was limited in the kinds of questions it could answer, especially research-heavy questions. This was because ChatGPT could not integrate data sources from the library. After Lappalainen and Narayanan tested and reviewed approximately 500 unique questions and answers, they saw that "the bot often generates non-existent links" and "the bot cannot answer questions that require real-time data or access to a specific resource" but noted that they expected that updating the data would be straightforward. Since Aisha was still in its infancy, this study did not fully assess its impact or capabilities. By licensing the API directly through OpenAI, Zayed University paid OpenAI for each query. It was anticipated that therefore, continued utilization of this tool could get expensive quickly if the volume of questions increased dramatically. This was in contrast to a vendor-supplied model, which charges a flat licensing rate that can be shared with other departments.[22]

Methodology

The authors utilized case study methodology to analyze the chatbot's performance during the Spring 2023 semester, which ran from January 25 through May 23. The authors' main objective was to determine how successfully the chatbot answered users' questions and where it struggled. Struggles were defined as the topics users were asking the chatbot about most and which seemed to stump it. To obtain the transcripts of the questions, the authors used the Ivy chatbot's export function to generate a list of transcripts and relevant metadata from the Spring 2023 semester. Then, transcripts with no interactions with the chatbot, (a person clicked the chatbot but did not continue with their interaction) were filtered out. This produced a random sample of 816 transcripts out of 2,504. Ivy's software creates a unique identifier for each transcript, but, for ease of use, each transcript received during the semester was numbered from 1 to 816. The authors used the Calculator Soup online random number generator to select a number of transcripts as a representative sample, wanting a range of a minimum of 1 and a maximum of 816, not allowing for repeats.[23] The authors chose 101 with no tickets as one dataset, which represented 12 percent of the total 816, with a margin of error of +/-7.27 percent. The transcripts chosen included those specified by the generator, which was designed to ensure randomness, and the total of 101 was chosen because it is slightly above the recommended 10 percent sample size, which would have been eighty-one.

It is also important to note that, normally, the library chatbot received questions users asked to the chatbot on the library or Blackboard site. However, during the transcript review process, the authors found that IT-related questions were being directed to the library instead of IT. This was due to IT mistakenly including a copy of the library's chatbot instance on an IT webpage, instead of the IT chatbot, which would normally be present. After discussion about what would be the acceptable amount of IT data that would influence the results, the authors removed eighteen transcripts from the random sample and removed four from the purposive sample. For the eighteen removed from the random sample, the authors replaced them with another set of eighteen randomized transcripts. There were no additional replacements to the purposive sample, since no replacement transcripts were available.

The second dataset consisted of all 39 chatbot tickets received during the Spring 2023 semester. Each ticket was created when the chatbot was stumped by a user's question, and the chatbot invited the user to fill out a ticket form. When users filled out this form, these tickets were sent to the authors along with the user's name and email address. The authors were then able to respond directly to each ticket. Transcripts selected in the online number generator that also had tickets attached to them were excluded from the 101 and set aside for this separate purposive sample. These were reviewed because the authors presumed that a user filling out a ticket meant that they were dissatisfied with the chatbot's answers and were seeking further assistance. Both the random and purposive samples were scrubbed of identifying factors such as IP addresses, users' names, and email addresses. IRB exemption was granted based on the use of retrospective data and lack of direct human interaction.

Rubric Score Creation

Once the data was ready for analysis, the authors created a rubric to score each transcript and ticket for accuracy, inaccuracy, completeness, and incompleteness. The authors turned to literature focused on scoring library reference questions and virtual chat interactions to understand what components should be present in the rubric. These scores would later be tied to subject areas through a content analysis, with the ultimate goal of determining the major subject areas that were difficult for the chatbot to respond to correctly. During Spring 2024, the two datasets, the random sample of 101 transcripts and 39 tickets, were analyzed and scored using the rubric.

Incorporating RUSA Guidelines

Reference librarians have often used the RUSA Guidelines for guidance on how to interact with patrons.[24] The Ivy chatbot has some of the questions RUSA advises librarians to ask already built into its software. For instance, the chatbot greets the patron, asking how it can be of assistance. It requests further information and lets the patron know that it is available as additional help is needed. The authors looked for further indications that the chatbot was using RUSA guidelines as a means of determining whether the chatbot was attempting to communicate its ability to successfully assist the patron. This was done in order to evaluate the answers the chatbot provides to users. Within the Ivy-supplied metrics, the authors noted the functionary response of the chatbot when it struggled to answer user questions. This was known as "low confidence." Examples of low confidence responses included "I don't know how to respond," or "I found several topics that may answer the question," or "What page are you looking for specifically?" When the chatbot could not answer a user's question at all—labeled as "no confidence" by Ivy—the chatbot would tell the user to "please fill out the form below and we'll get back to you soon. To speak to a librarian, click the message below."[25] Noting the similarities between the chatbot's low confidence or no confidence responses and RUSA's engagement criteria, with suggested questions like "Could you please tell me a little more about what you are looking for?" and "Are you looking for particular types of sources, such as books, websites, etc.?," the authors decided to include the presence of chatbot low confidence and no confidence responses as competency scales in the rubric score.

Ivy-Supplied Metrics

Only transcripts with no confidence responses that also did not have tickets filled out were included in the random sample. The next competency scale was what the authors called tickets refused, in which the user cancelled the no confidence response and chose to continue asking the chatbot questions. This scale was important for understanding users' behavior as they continued their conversations with the chatbot. The authors also incorporated an Ivy-supplied metric for classifying transcripts; the system prompted users to select a thumbs-up or thumbs-down icon (see Table 1). This icon choice appeared to users when asked to rate individual responses from the chatbot. This competency scale was important because it potentially separated out aspects of the chatbot's responses that were correct and indicated that a patron found the chatbot's answer satisfactory, even if all answers provided to the user were not. Separating the perceived good responses from the bad became a common aspect of evaluating transcripts because users often asked multiple questions on different topics or more importantly, rephrased their follow-up questions when the chatbot struggled to answer them. The four competency scales (low confidence, no confidence, ticket refused, thumbs up, thumbs down) comprised the basis for the rubric. The authors then developed user variables to complete the rubric.

User Behavior Variables

User Satisfaction

When assessing in-person reference, Joan C. Durrance tested librarians' interpersonal skills. These included interviewing and listening skills and the effectiveness of librarian approaches to user questions. This testing also included "search strategy, accuracy, and the ability to provide the questioner with a satisfactory answer." Durrance's study revealed that while accuracy was important to users, interpersonal skills played a large factor in their willingness to return and ask a question later. The willingness to return

Table 1.

Chatbot competence scale

Score	Description	Criteria
1	Low Confidence Bot Responses	Bot Responses included "I don't know how to respond" or "I found several topics that may answer the question" or "What page are you looking for specifically?"
2	No Confidence Bot Responses	Bot asked users to fill out a form to create a ticket or speak to a live librarian.
3	Ticket refused	Bot offered to create a ticket, but user declined to or started but cancelled creating a ticket.
4	Thumbs Down	User rated the question with a thumbs down (negative) rating
5	Thumbs Up	User rated the question with a thumbs up (positive) rating

"permits the examination of accuracy within the context of other factors that contribute both to the success of the reference interview."[26] For example, users often rate their satisfaction with a reference interaction as "equal in importance to the accuracy of the information provided in the evaluation of the reference service."[27]

In order to better understand the importance of user satisfaction, the authors of this study searched for variables that would measure this component (see Table 2). The first variable identified was whether a user explicitly expressed satisfaction with the chatbot interaction, using phrases such as "it works! thanks!" [sic]. Conversely, the second variable concerned instances when users explicitly expressed dissatisfaction with the chatbot interaction through phrases that indicated criticism such as, "that's not what I want." These examples were derived from chat reference evaluation studies that assessed dialog, emotions expressed through punctuation, and closings that could be interpreted as satisfaction or frustration.[28] In a similar vein, the authors identified a third user variable related to whether the user explicitly asked for an agent, representative or librarian, which signified that the chatbot was not satisfying their informational needs. A fourth variable noted whether the user disconnected from the chat prematurely, before the question was answered or after the answer was given and the user provided no response.

A total of four of the randomized no ticket transcripts expressed satisfaction by saying "thank you." However, given that thank you may be a courtesy response, the authors examined the transcripts in more detail to see whether each user added an accompanying show of emotion or description of how they were helped. The authors determined that only three out of 101 transcripts noted satisfaction with an expression of thanks and positive emotion.

Table 2.
Study variables

Variable	Description
1	Chatbot's answer contains one or more keywords that match the user's answers
2	User disconnected from chat prematurely (before question was answered/after answer given and no response)
3	User explicitly states satisfaction (Thank you, have a good day)
4	User explicitly states dissatisfaction (That's not what I want)
5	User asks for agent, representative, or librarian
6	Was the question answered? Yes or No

For the other variable of premature disconnection, 93 percent of transcripts revealed premature disconnection from the chat, which may indicate that this was typical patron behavior when engaging with a chatbot, a finding that mirrored other studies that showed users often disconnect prematurely from online chat interactions.[29] The authors then constructed variables to measure the accuracy of the chatbot's answers based on response content and librarian review.

Accuracy and Completeness

Jeffrey Pomerantz noted that while a user's satisfaction with a reference interaction can be important, it does little for them if the information is incomplete or inaccurate, and users may not realize any potential inaccuracies. N.J. Belkin, R.N. Oddy, and H.M. Brooks echoed this when they suggested that a user who asks a reference question probably does not know enough to evaluate the answer for completeness or correctness. In other words, if they did know enough to make this evaluation, there would be little reason for them to have asked the question in the first place.[30] Based on this important distinction, the authors added two more variables: the presence of one or more keywords that matched those found within the author's question and whether the question was answered by the chatbot with a yes or no answer. This practice was discussed when evaluating chat transcripts as a gauge for accuracy.[31]

Breakdown of Competence Scales and Variables

Based on these variables and competence scales, the authors created a four-point rubric: 0-Complete and Correct, 1-Incomplete but Correct, 2-Incorrect, 3-Incomplete and Incorrect, 4-No Answer Provided (see Table 3). Each transcript received a point score and the number of transcripts with each score was calculated to determine how many transcripts

Table 3.
Scoring rubric

Score	Meaning	Criteria
0	Complete and correct	Must include Competence Scale 5, Variable 1, AND Variable 3, AND a Yes from Variable 6
1	Incomplete but correct	Must include at least one competence scale from Competence Scale 5 OR Variable 1, OR Variable 3, AND a Yes from Variable 6
2	Incorrect but complete	Must include either Competence Scale, 1,2, or 3, OR Variable 4 or 5, AND a No from Variable 6
3	Incorrect and incomplete	Must include either Competence Scale 1, 2, or 3 AND Variable 4 or 5, AND a No from Variable 6
4	No answer provided	No answer is given at all by the chatbot

were correct, incorrect, complete, or incomplete. To determine the weight of the rubric, the authors placed more emphasis on the variable measuring whether the question was answered correctly. This was based on earlier discussions from the literature about users not always knowing what the right answer was. More emphasis was also placed on this variable because many patrons did not express satisfaction or dissatisfaction with their chatbot encounters and exited the chat prematurely. To the authors, this suggested that determining whether the question was answered correctly or incorrectly was a more reliable measure.

For all transcripts with tickets, the authors awarded only scores of 2 and 3 based on the criteria previously described. One author scored all tickets with a 3, while the other rated 38 out of 39 with a 3 rating and one transcript with a 2. The difference in this case was based on the authors' judgment of whether the question was answered or not, which they disagreed about in this case.

The rubric scores determined the percentages of incorrectly answered questions and severity of their level of failure. This was essential for tying the percentages to the categories of questions that failed.

Content Analysis and Codebook

The authors employed inductive coding of the samples' questions to determine content categories such as library hours, textbooks, or financial aid. The frequency of each subject code tied to the transcript rubric scores revealed question categories the chatbot struggled to answer correctly and completely.

Each author conducted an independent review of the data in order to ensure a high inter-reliability rate. The authors then met to reconcile the codes, as well as the 39 transcripts where users submitted tickets. At this stage, a total of 37 codes were created, and each transcript was assigned one code which described the content. The full codebook can be found in Appendix A.

It is important to note that Ivy transitioned to the ChatGPT API with the launch of IvyQuantum on April 3, 2023. Thus, for approximately the first half of the semester, the chatbot used its original rules-based model, and for the second half, its functionality included both rules-based responses and generative responses powered by ChatGPT. To better understand the effect of ChatGPT's API, the authors also attempted to see if there was any improvement in answers post-April 3 as a result.

Normally, the library chatbot received questions users asked on the library or Blackboard site. However, during the transcript review process, the authors found that IT-related questions were being directed to the library instead of IT. This was due to IT mistakenly including a copy of the library's chatbot instance on an IT webpage, instead of the IT chatbot, which would normally be present. After discussion about what would be the acceptable amount of IT data that would influence the results, the authors removed eighteen transcripts from the random sample and removed four from the purposive sample. For the eighteen removed from the random sample, the authors replaced them with another set of eighteen randomized transcripts. There were no additional replacements to the purposive sample, since no replacement transcripts were available.

Results

Success Rate

Figure 1 represents the rubric-scored distribution of the random sample of 101 transcripts. A 39 percent success rate for transcripts with rubric scores of one indicated that answers given were correct but incomplete. This showed the chatbot was succeeding in answering the questions in general but not at the highest satisfaction scores possible on the rubric. This was also evident in the 42 percent of transcripts that received rubric scores of two, in which answers were incorrect but complete. Comparing these rates to the widely cited 55 percent librarian success rate in answering reference questions, one could say the chatbot was performing rather well in comparison, especially given it is not human, let alone a librarian.[32]

. . . the chatbot was succeeding in answering the questions in general but not at the highest satisfaction scores possible on the rubric.

Codes

Figure 2 represents the purposive sample of 39 transcripts where users filled out help tickets. The category of "others" includes miscellaneous categories—ones where the subject users asked about was found to have only occurred once in the sample.

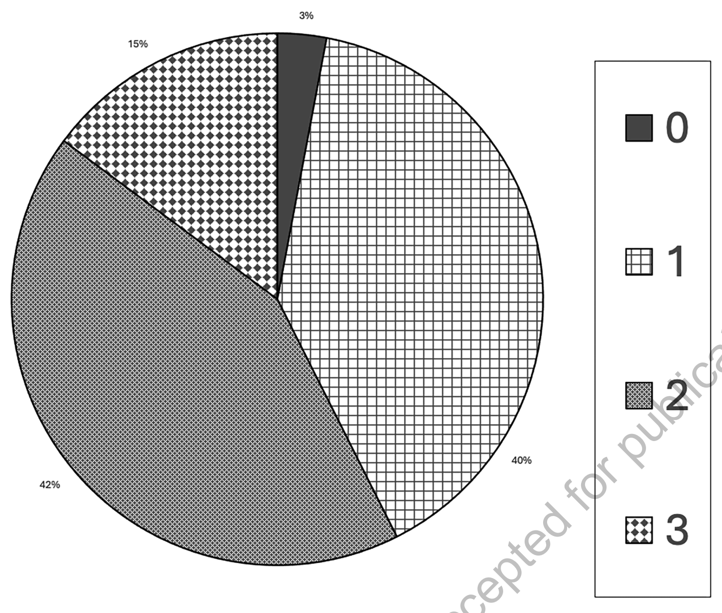

Figure 1. Rubric score distribution for the random sample with no associated help tickets.

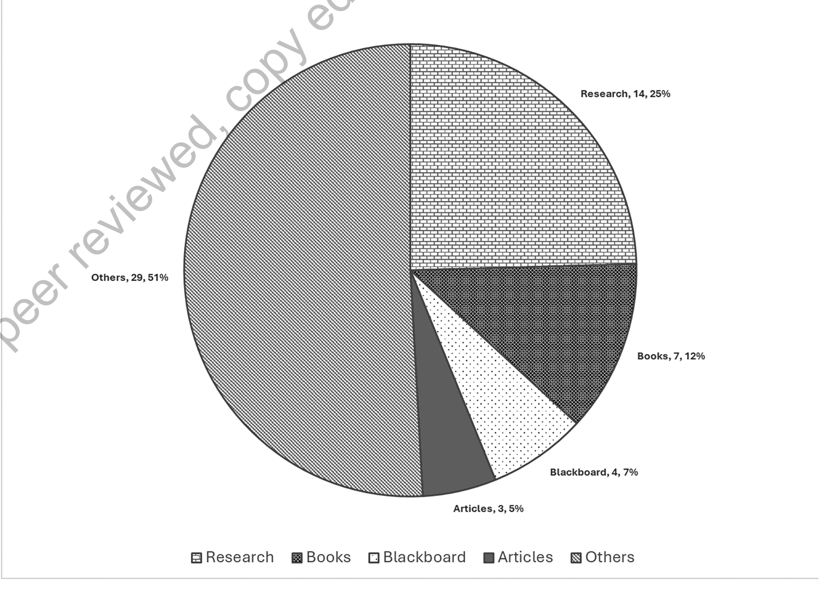

Figure 2. Rubric score distribution for the purposive sample.

The dominance of research- and book-related tickets aligned with the author's previous findings, which indicated that the most common topics users filled out tickets for textbooks, research, and off-campus access.[33]

Books, Articles and Requests for Agents

Books was the dominant code, accounting for 14 percent of all transcripts with rubric scores of two (where the answer supplied by the chatbot was complete but incorrect). This was closely followed by the 12 percent with the code "articles" and then the miscellaneous "other" category. In the purposive sample, the code with greatest frequency was research, followed by books. For rubric scores of three, where answers were incorrect and incomplete, "requests to speak with an agent" was the most dominant code (see Figure 3).

ChatGPT's Impact

Of the thirty-nine questions with help tickets, ten questions were asked before the ChatGPT integration (.2 questions per day), and twenty-nine questions were received afterward (.58 per day.) In other words, questions with help tickets increased after the ChatGPT launch.

Figure 4 demonstrates the performance of the chatbot prior to April 4, 2023 and afterward when Ivy added integration of ChatGPT's API and allowed indexing of the entire library website. Sixty-five questions without tickets were received prior to the ChatGPT integration, and 36 afterwards. This represents a 45 percent drop in questions received. During this same period, questions with tickets grew from 10 to 29, representing a 190 percent increase.

In terms of questions without tickets, the ratios of ratings also changed. Prior to Ivy's ChatGPT integration, 0s accounted for 3 percent of answers, 1s for 34 percent, 2s at 45 percent, and 3s at 18 percent. After IvyQuantum went live, this changed to 3 percent with a score of 0, 50 percent scored 1, 39 percent 2s, and 8 percent 3s. This means that 0-scores stayed relatively similar, 1-scores increased as a percentage, 2-scores decreased, and 3-scores decreased as well.

Most notably, the number of transcripts that received a score of 1 increased significantly, with 34 percent of questions prior to the GPT implementation receiving this score and 14 afterward (a 47 percent increase).

Discussion

The results of the assessment highlighted three main question areas the chatbot struggled to answer successfully: requests for an agent, requests for books, and requests for articles and research help. In the purposive sample, research requests were the most popular questions users submitted tickets for because they were

The results of the assessment highlighted three main question areas the chatbot struggled to answer successfully: requests for an agent, requests for books, and requests for articles and research help.

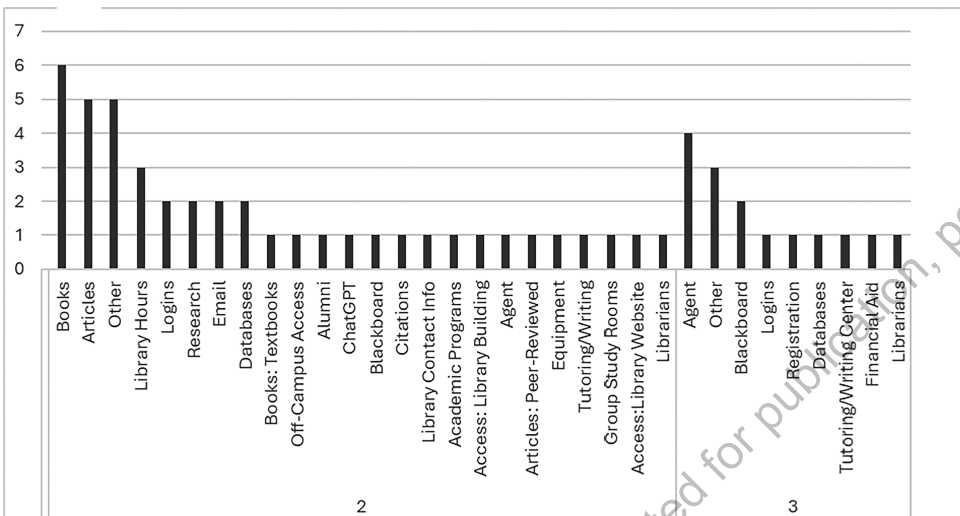

Figure 3. Frequency of codes for rubric scores two and three for the random sample.

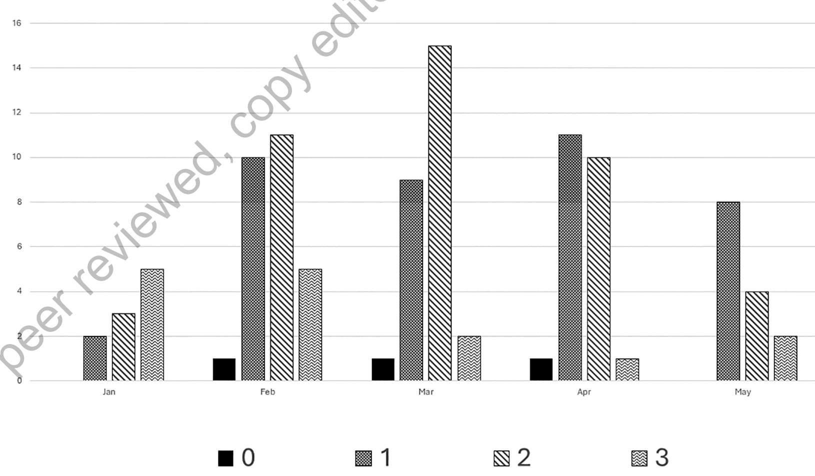

Figure 4. Rubric scores assigned during the assessment period.

unhappy with the chatbot's responses. Next, the authors will discuss changes that were made to address the uncovered weaknesses in the chatbot's performance.

Requests for an Agent

Based on the results, the question the chatbot struggled to answer the most were requests for an agent, which, unlike other university departments such as IT, Financial Aid, and Admissions, the library chose not to set up. When offered a librarian via library chat as an alternative to an agent, users still insisted on speaking to an agent instead. This perhaps demonstrated a lack of awareness among patrons of how a librarian could provide similar assistance to what another department's agent might. It also showed a cross-over behavior that users may have come to expect between chatbots from different departments, something the authors had not considered when thinking about live agents.

To address this question, the authors could consider enabling the agent feature, mimicking other departments in terms of available times. However, since the authors also staff the reference desk and live chat service at the library, using the agent feature may not be sustainable, depending on how many chats are received on a daily and weekly basis, where an agent might be requested. Alternatively, the authors could review the customized response to the request for the agent question to see whether changes might be made to direct users more easily to the library chat service or to the ticket feature.

The current chatbot response to request for an agent from a user is, "You can chat with a librarian 24/7 on [our live chat site]. We generally respond to emailed questions within 48 hours when the library is open. You can email us [here]. You can reach us by phone during Library hours at [number]. You can also stop by the Reference Desk on the first floor and talk with a Librarian during Library hours."[34] This could be trimmed to just the first line and "librarian" replaced with "library agent" so that the response would read, "You can chat with a library agent 24/7 on our live chat site. May we connect you to a live library agent?" User testing in this instance may be beneficial to determining whether a change in verbiage would result in a change in behavior.

Requests for Books

The chatbot struggled to correctly answer questions about books, both in the random and purposive samples. This was consistent with the earlier discussed assessment of the chatbot's implementation in 2019 and with University of California Irvine's study, which also noted books as a very popular question category. The library's virtual chat received 431 questions about textbooks during the Fall 2023 semester. Because of their frequency, the authors of this paper worked to address book-related questions using the chatbot. Since book-related information is fairly constant and is not nuanced, the authors' belief was that the chatbot could easily retrieve textbook information by crawling the library website.

The chatbot struggled to correctly answer questions about books . . .

Creating a Reserve Textbooks Webpage

The transcripts showed that certain books were found via the library's LibGuides, meaning the chatbot was using this data to generate its answers. To make the reserve textbooks data available on the website, the authors worked with their library's access services department to generate a report in Google Sheets, which included all books currently on reserve. Access Services provided the report with columns for title, author, edition, call number, and hyperlinked URL, pulled from book records in the library's Primo VE discovery system.

One of the authors then created a new library webpage based on the Google Sheets document, which included 1898 titles, for the chatbot to crawl. However, initial testing revealed the chatbot did not consistently give correct responses. Guided by the literature, the authors concluded that the chatbot was likely hallucinating due to the size of the database. Ivy support advised the authors to create a CSV file from the reserve textbooks list and share it with them to ingest directly into the chatbot knowledge base, instead of relying solely on crawling the webpage. Ivy speculated that the structured data in the CSV would address the chatbot's hallucinations because they occurred most frequently in the URL field, which linked to the book's real-time availability. Ivy speculated this was a result of the chatbot's predictive behavior, which caused URLs to be random strings of text, instead of a structured data field associated with other fields. Without a table indicating that each column was a field of structured data, and not simply text, the chatbot struggled to accurately supply URLs.

Without a table indicating that each column was a field of structured data, and not simply text, the chatbot struggled to accurately supply URLs.

After the CSV file was added to the knowledge base, the authors tested the chatbot using questions about the textbooks and found its ability to respond with accurate information successful. Figure 5 shows the chatbot's response to a textbook question once the reserve textbooks list was ingested into the chatbot's knowledge base.

The other librarians at Lehman College also appreciated the new reserve books list and webpage. They used it to support in-person reference interactions, as a backup tool for locating textbooks when searching the library's discovery system and as a way of confirming these items were held as reserve materials. Between its deployment on January 14, 2024, and August, 2024, the web page has been viewed over 600 times by 346 distinct users. The authors plan to update this list for future semesters.

Articles and Research Requests

The authors noticed that requests for articles were often tied to research requests in the random sample and in the purposive sample, in which research requests were the most coded category. IvyQuantum struggled with understanding nuance and referring to sources when responding research questions in the same way that Lai's study noted that ChatGPT did. The authors plan to redo the chatbot's responses to such requests, which are currently: "Find useful articles, books and journals in our online database

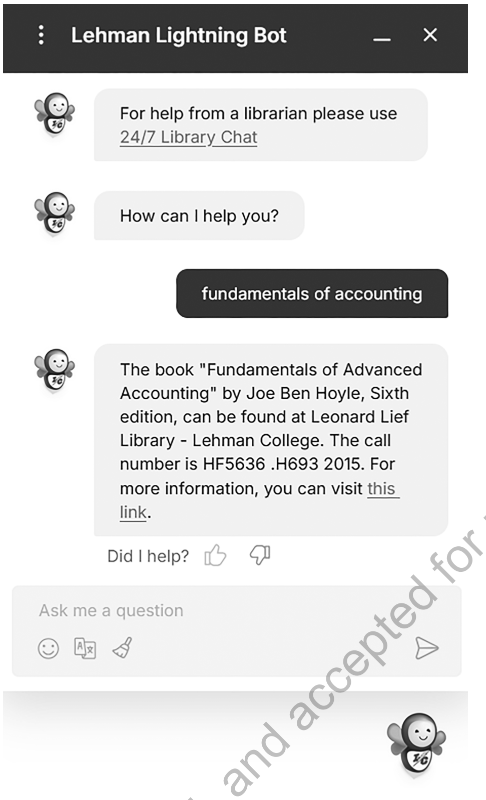

Figure 5. Example chatbot transcript.

by clicking here!" or the no-confidence response of: "Please fill out the form below and we'll get back to you soon. To speak to a librarian, click the message below." Possible solutions include hyperlinking to virtual chat before offering the option of filling in the form or inserting the form along with links to the library databases, with messaging such as "for research help finding articles, chat with a librarian."

Creating a Chatbot Landing Page

The library website lacks a dedicated page describing the chatbot and how it is intended to be used. The University of Oklahoma's Bizzy Chatbot and San Jose State University's Kingbot include pages like this.[35] Creating a landing page that links to the chatbot could clarify that the chatbot is not intended for research questions and direct users to library chat. A separate webpage could also function as a second chatbot access point, which, combined with its description and purpose, could further clarify for users what the chatbot is best suited for.

The ChatGPT API

ChatGPT's API in IvyQuantum has had major effects on the chatbot's ability to answer users' questions, thanks to its crawling the library website daily, its FAQ, and research guides. This means the chatbot can retrieve virtually any information it can match with the user's request. It also incentivized the library to keep its website up-to-date and create new content based on users' questions. The authors receive daily emails about how much content has been added, removed, and updated from Ivy, which is helpful for checking that the knowledge base is current. New data can be submitted to IvyQuantum's knowledge base in multiple forms, such as calendar feeds, documents and pdf's, data_tables, and custom responses. This offers flexibility for data sources to be integrated on the website and options for the authors to create new ones.

. . . the chatbot can retrieve virtually any information it can match with the user's request. It also incentivized the library to keep its website up-to-date and create new content based on users' questions.

IvyQuantum enables editing of chatbot interactions after they occur, which allows easier daily management and real-time updates. The ChatGPT integration also promotes more generative answers rather than retrieval-based ones, which should improve the accuracy rate and require less maintenance on the authors' parts. For example, at the time of the writing of this article, the authors were able to upgrade the default no confidence response to use a generative response rather than a retrieval-based one. While no data is yet available to see how patrons respond, the authors expect this change to improve the chatbot's ability to answer the user's request beyond just offering them an option to create a ticket. The authors also plan to perform a custom response audit to see which ones are most effective at answering user's questions and delete those that are not, so that the chatbot can move toward generative responses in their place.

Ivy customers are limited in how many data sources they can employ based on their subscription model, a disadvantage in comparison to using ChatGPT and all of its amassed data on its own. The authors' library does not pay for Ivy but offers a free version of ChatGPT. Libraries looking to subscribe to IvyQuantum or a similar product may be able to have their college cover the cost, especially if they share a chatbot like the Leonard Lief library does with Blackboard. They can also license ChatGPT's API directly like Zayed University, where prices vary based on model and tokens, which dictates the number of queries that can be received and their complexity level.

Best Practices for Chatbot Maintenance and Performance

For the authors, this assessment demonstrated the importance of administrators of chatbots like IvyQuantum reviewing users' queries and finetuning the chatbot's answers to ensure they are correct or, at the very least, complete enough to guide users to the next step and encourage them to return. Based on this article's findings, the authors highly recommend regular reviews of the chatbot's transcripts to allow administrators to make changes to custom responses, the knowledge base, and the library website to improve chatbot answers. While the authors will likely take on a monthly transcript review, the IT department at the college does a daily review, while other departments' review schedule is unclear. Best practice would dictate a review that accounts for volume, accuracy rate, and administrator's schedule.

Creation and use of the rubric and codebook has reinforced a need for understanding what users are asking chatbots like IvyQuantum, ChatGPT, and others about, and how well each chatbot is doing at answering them. While this study's rubric was created specifically for evaluation of IvyQuantum, the authors hope others can adapt and modify it for other vendor-managed chatbots and ChatGPT. The rubric's emphasis on accuracy, which the literature shows is ChatGPT's biggest weakness, proves particularly useful. Librarians know that students are using ChatGPT for research and course assignments and should be aware of its limitations in providing consistently accurate information. Libraries looking to incorporate ChatGPT or other chatbots into their library sessions can utilize the information gathered in this article about managing and improving chatbots vis a vis the library website. This will help to ensure the product's standing as an asset to virtual services, such as library chat.

. . . the authors highly recommend regular reviews of the chatbot's transcripts to allow administrators to make changes to custom responses, the knowledge base, and the library website to improve chatbot answers.

Libraries looking to implement a chatbot are advised to work with their college or university to select the one that best suits their needs in terms of cost, maintenance, and control. Having designated librarians committed to the upkeep of the chatbot will make a significant difference in its performance. For a chatbot to be successful, buy-in from others in the library is also essential so offering demonstrations and product testing opportunities is necessary. This will help ensure the chatbot's promotion and use and demonstrate the return on investment in the chatbot as a library website enhancement.

Limitations

The data in this study came from the chatbot on the library page and the Blackboard page, and the library could not separate those queries intended for assistance with Blackboard from those intended for the library. The study sample contained data from only one semester, during which ChatGPT's functionality was added to Ivy's product, so the chatbot could not effectively be compared to previous semesters. In other words, the data assessed two types of technology, something which did not happen during previous semesters. A full semester of IvyQuantum may have demonstrated more of the API's effect on responses to user questions.

Future Research

Future research might include an automated analysis of transcript content using text-analysis tools or using tools like sentiment analysis to better understand patron feelings about the chatbot. Research could also be compared year-to-year and contrasted with other libraries who have licensed the ChatGPT API directly. A partnership with another library that uses IvyQuantum could also be explored to conduct research comparing users' questions or sharing best practices for chatbot setup and maintenance. Long term, the authors hope to measure the chatbot's success in answering basic reference questions such as library hours, looking up books, and finding databases, as a means of freeing up library virtual chat for more in-depth research questions.

Conclusion

This research has demonstrated the importance of formal assessment when implementing and managing a chatbot and the value of using a rubric in the process. This project highlighted some advantages of using vendor-managed chatbots like IvyQuantum, along with the ChatGPT API in contrast to using ChatGPT on its own or just licensing the API. The chatbot struggled to answer questions about library agents, books, and research and required upkeep to ensure accuracy and completeness of its responses. Maintaining the library website and all of its data is vital to ensuring the chatbot can crawl the most accurate and complete data. Future research is needed to further explore the advantages and disadvantages of chatbots and develop best practices for deploying chatbots on library websites. As artificial intelligence continues to evolve, libraries with chatbots are well-equipped to harness this technology to meet patrons' needs in a 24/7 information-seeking environment.

Acknowledgement

The authors would like to thank Kristine Brancolini and Marie Kennedy of the Institute for Research and Design in Libraries (IRDL) for their guidance in developing this project.

Michelle Ehrenpreis is the electronic resources librarian and assistant professor at Lehman College, CUNY, email: michelle.ehrenpreis@lehman.cuny.edu: ORCID: 0000-0003-0238-9921.

John DeLooper is the web services-online learning librarian and assistant professor at Lehman College, CUNY, email: john.delooper@lehman.cuny.edu, ORCID: 0000-0003-3965-9251.

Appendix

Coding Book

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Academic Programs	Non-library related questions	Patron has questions about academic programs and their offerings at Lehman	Questions about Lehman's academic programs	Questions about registering for classes	physician assistant program
Access: Library Website	Library website	Patron has questions accessing the library website	Patron is trying to or cannot access library website	Patron is trying to access a database or something else	i am having trouble searching Libraries > Problems > Access > Codes

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Access: Library Building	Library building	Gaining access to the library building	Questions about accessing the library building	Accessing the library website or library databases	I am a current Lehman graduate student and I want to go to the library Sunday. May I bring my wife onto the campus?
Advising	Non-library related questions	Patron is trying to get information about advising, course prerequisites, etc.	Patron is trying to reach advising office, has questions about what classes are required	Patron has already graduated and needs transcripts; patron needs to register for classes	How would my student find out her academic advisor and contact info?
Agent	Ivy Live Agent Feature	patron is trying to reach a real person (live agent) to get information	patron asks for agent, live agent, or representative	when patron requests to speak to a librarian, or somebody specific from another office	I need to speak to a representative, I need to speak to agent
Alumni	Alumni	Questions from alumni using the library	Questions from or about alumni	Questions about students accessing the library	what is the benefit for Lehman Alumni?

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Articles	Library materials: articles	Articles published in journals, magazines, or newspapers	To describe questions where users are looking for articles from the aforementioned areas	Users asking research questions that may include gaining access to these materials	i want to find journal articles
articles: peer-reviewed	Library materials: peer-reviewed articles	Articles that are published in peer-reviewed academic journals	Peer-reviewed journal articles	Articles from newspapers or magazines; non-peer reviewed articles	how can i view peer reviewed articles
Blackboard	technology	Questions related to the Blackboard Learning Management System (LMS) used by Lehman for classes	questions about access and logging into Blackboard	questions about logging into any other college systems	Hi I can't log into my blackboard
Books	library materials: books	questions related to print books in all forms, whether circulating or reserve, or those that can be checked out or used within the library	to describe chatbot transcripts/tickets where users are looking for books on research topics that they have been assigned.	users looking for eBooks or textbooks	I'm looking for a math book

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Books: Textbooks	Library materials: textbooks	Questions asking for specific textbooks or looking for general textbooks	Questions asked about specific textbook titles or finding general textbooks	Questions about the bookstore or selling books	I need a textbook name teach yourself how to learn strategies you can use to ace any course at any level by my professor
ChatGPT	Chatbots	Questions about ChatGPT or other related chatbots	Patrons who ask questions about chatbots, how they work, other chatbots, especially ChatGPT	Questions about Ivy specifically	Good bot. (I called your chat bot a good bot and it broke. Ignore this ticket and close it and enjoy the rest of your shift :))
Citations	Citations	Citations questions about formatting, style	Questions about citations and formatting/style guides	Questions about writing papers	citation machine asa
Classes	Non-library related questions	Questions about location and times of specific college classes	To describe questions about class times and location	Questions about registering for classes	What time does a biology class start?
Databases	Databases	How to find all the library databases or specific ones	Questions asked about finding databases or accessing specific ones	Questions about off-campus access to databases	How can I get to databases to use the oxford dictionary?

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Email	technology	accessing email, resetting password to email etc.	questions about accessing, logging in and resetting email passwords	questions about logging into other college systems	How to log into outlook
Equipment	Library equipment	Requests for information related to borrowing laptops, calculators, iPads, headphones, etc.	Questions about circulating any electronics	Questions about classroom equipment, questions about borrowing laptops	I'd like to borrow a laptop
Financial Aid	Non-library related questions	Looking for information or assistance from the financial aid office	To describe chatbot transcripts/tickets where users are asking for assistance finding information related to financial aid	Where users are asking for live agent or other department	i need to talk to financial aid
Graduation	Non-library related questions	Questions pertaining to graduation processes, procedures, and information	Questions about graduation, when it happens, what are its requirements, etc.	Questions from people who previously graduated and need transcripts or course verifications	I graduate from lehman I think on 2006 but I need information about when I started and when I graduate

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Group Study Rooms	Group study rooms	Booking access and availability of group study rooms	Questions about booking access and availability of group study rooms	Questions about accessing the library	are there any 1 person study rooms?
IT Center: Help Desk	Technology	Questions pertaining to reaching the College Help Desk for IT Assistance	questions about accessing, speaking, and finding the IT Center Help Desk	questions about contacting IT about logins	can you print at the it center
Librarians	Librarians	Patron needs to talk to a librarian	Patron is trying to reach a librarian	Patron is trying to reach an agent	Subject > Contact information > Librarians
Library Contact Info	Contacting the library	Contact Information for the Library	Questions about reaching departments in the library	Trying to reach a librarian or agent	Hi, what is the e-mail for the Lehman Library
Library Hours	Library Hours	the hours that the library building is open to the campus	to describe chatbot transcripts/tickets where hours of the library are asked about and where to find them on the library website	users looking for hours of the chatbot or library chat or other departments hours	Library is open tomorrow?
Logins	Technology	Patron has question about login processes/procedures	Patron has login questions	When patron needs a password reset	Lehman 360 help

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Off-Campus Access	Access to library materials off-campus	How to access library resources from off-campus	To describe chatbot transcripts/tickets where users are asking about how to access resources off-campus or having trouble doing that	Users asking about other logins like Lehman 360 etc.	why I dont hace access to the databases, is there a problem going on
Other	Does not fit in with other categories	Any areas where nothing is present	When no other category fits better	When another category fits better	t-shirts question
Password Reset	Technology	Patron needs to reset a password	Patron asks for a password reset	When patron specifies which password needs to be reset, i.e. CUNYFirst, Lehman360, etc.	Hi I need to reset my Lehman email password
Registrar	Non-library related questions	Information about student records, choosing courses, etc.	Students asking questions about transcripts, choosing classes	When students need to register for classes	INC Grade: What Does It Mean

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Registration	Non-library related questions	General questions about signing up for classes	Questions about registering for classes	When patron needs to talk to their advisor and requests advisor info	hi I need to communicate with registration department
Research	Research using the library's resources	Questions related to research using the library's resources	To describe chatbot/ transcripts tickets where users are asking research-related questions that require more in-depth answers	Users looking for exact titles of books or other materials	Hi, I am looking for resources on the Civil Rights movement based in Detroit, Chicago and NYC.
Research Guide	Research	Question about library research guides	Questions about research guides, when to use research guides	When a patron needs help with research	How to Find the Research Guide
Student ID	Technology	Questions related to student IDs	Questions about getting a student ID, using a student ID, Virtual ID, ID cards vs. ID numbers, etc.	Questions about logins to library systems or website	I need my student id number

Code Label	Theme	Full Definition	When to Use:	When not to use:	Example:
Tutoring/Writing	Non-library related questions	Questions related to getting tutoring services or writing services, or those related to the Academic Support Services Department	When people ask for information about tutoring, tutoring services, or to get in-person or online tutoring and writing help	When students asking for help from the library	Tutoring Center

Notes

1. Edward Winstead, "Can AI Chatbots Correctly Answer Questions about Cancer?" *Cancer Currents* (blog), October 3, 2023, https://www.cancer.gov/news-events/cancer-currents-blog/2023/chatbots-answer-cancer-questions.
2. "Introducing ChatGPT," OpenAI, November 30, 2022, https://openai.com/index/chatgpt/.
3. Ethan Mollick, "ChatGPT Is a Tipping Point for AI," *Harvard Business Review*, December 14, 2022, https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai.
4. Bergur Thormundsson, "Work Tasks Completed Using ChatGPT U.S. 2023," Statista, April 20, 2023, https://www.statista.com/statistics/1378998/chatgpt-use-tasks-us-by-type/.
5. Lehman College, "About Lehman College," Lehman College, 2024, https://www.lehman.edu/about/.
6. "Ivy.Ai Homepage," Ivy.ai, 2024, https://ivy.ai/.
7. Michelle Ehrenpreis and John DeLooper, "Implementing a Chatbot on a Library Website," *Journal of Web Librarianship* 16, no. 2 (2022): 120–42, https://doi.org/10.1080/19322909.2022.2060893.
8. Ibid.
9. Ibid.
10. Ibid.
11. Ibid.
12. Ibid.
13. "Ivy.Ai Scales GenAI-Powered Communication with Zilliz Cloud Vector Database," 2024, https://zilliz.com/customers/ivy.ai.
14. "What Is RAG? - Retrieval-Augmented Generation AI Explained," Amazon Web Services, Inc., 2024, https://aws.amazon.com/what-is/retrieval-augmented-generation/; "What Is Retrieval Augmented Generation (RAG)?" Google Cloud, 2024, https://cloud.google.com/use-cases/retrieval-augmented-generation.
15. "What Is an API (Application Programming Interface)?" Amazon Web Services, Inc., accessed October 11, 2024, https://aws.amazon.com/what-is/api/; Michael Goodwin. "What Is an API (Application Programming Interface)?" IBM, April 9, 2024, https://www.ibm.com/topics/api.
16. Kate McAlpine. "Understanding Attention in Large Language Models," *Michigan Engineering News* (blog), December 13, 2023, https://news.engin.umich.edu/2023/12/understanding-attention-in-large-language-models/; Aditya Rajora "Chatbot Using Transformer from Scratch," *Medium* (blog), May 29, 2022, https://medium.com/@57fdaditya/chatbot-using-transformer-from-scratch-6ec45db2a208.
17. "What Is Generative AI?," McKinsey & Company, April 2, 2024, https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai.
18. "Rule-Based Chatbots: Rule-Based Chatbots Cheatsheet," Codecademy, 2024, https://www.codecademy.com/learn/rule-based-chatbots/modules/rule-based-chatbots/cheatsheet.
19. Sharesly Rodriguez and Christina Mune, "Uncoding Library Chatbots: Deploying a New Virtual Reference Tool at the San Jose State University Library," *Reference Services Review* 50, no. 3/4 (January 1, 2022): 392–405, https://doi.org/10.1108/RSR-05-2022-0020.
20. Ibid.
21. Katie Lai, "How Well Does ChatGPT Handle Reference Inquiries? An Analysis Based on Question Types and Question Complexities," *College & Research Libraries* 84, no. 6 (2023), https://crl.acrl.org/index.php/crl/article/view/26102/34024.
22. Yrjo Lappalainen and Nikesh Narayanan, "Aisha: A Custom AI Library Chatbot Using the ChatGPT API," *Journal of Web Librarianship* 17, no. 3 (July 3, 2023): 37–58, https://doi.org/10.1080/19322909.2023.2221477.
23. Edward Furey, "Calculator Soup - Online Calculators," CalculatorSoup, July 16, 2024, https://www.calculatorsoup.com/index.php.

24. “Guidelines & Resources,” American Library Association/Reference and User Services Association, 2023, https://www.ala.org/rusa/guidelines-resources; MARS Digital Reference Guidelines Ad Hoc Committee, “Guidelines for Implementing and Maintaining Virtual Reference Services,” *RUSQ: A Journal of Reference and User Experience* 50, no. 1, https://journals.ala.org/rusq/article/view/3168.
25. “Ivy - Admin Portal,” IvyAI, 2024, https://app.ivy.ai/admin/analytics/dashboards/bot.
26. Joan C. Durrance, “Reference Success: Does the 55 Percent Rule Tell the Whole Story?” *Library Journal* 114, no. 7 (April 15, 1989): 31–36.
27. Jeffrey Pomerantz, “A Conceptual Framework and Open Research Questions for Chat-Based Reference Service,” *Journal of the American Society for Information Science and Technology* 56, no. 12 (2005): 1288–1302, https://doi.org/10.1002/asi.20215.
28. Marie L. Radford, Lynn Silipigni Connaway, Patrick A. Confer, Susanna Sabolcsi-Boros, and Hannah Kwon, “‘Are We Getting Warmer?’ Query Clarification in Live Chat Virtual Reference,” *Reference & User Services Quarterly* 50, no. 3 (Spring 2011): 259–79, https://doi.org/10.5860/rusq.50n3.259; Samantha Harlow, “Beyond Reference Data: A Qualitative Analysis of Nursing Library Chats to Improve Research Health Science Services,” *Evidence Based Library and Information Practice* 16, no. 1 (March 15, 2021): 46–59, https://doi.org/10.18438/eblip29828.
29. Michael Mungin, “Stats Don’t Tell the Whole Story: Using Qualitative Data Analysis of Chat Reference Transcripts to Assess and Improve Services,” *Journal of Library & Information Services in Distance Learning* 11, no. 1–2 (January 2, 2017): 25–36, https://doi.org/10.1080/1533290X.2016.1223965.
30. N.J. Belkin, R.N. Oddy, and H.M. Brooks, “ASK for Information Retrieval: Part I. Background and Theory,” *Journal of Documentation* 38, no. 2 (January 1, 1982): 61–71, https://doi.org/10.1108/eb026722; Jeffrey Pomerantz, Lili Luo, and Charles R. McClure, “Peer Review of Chat Reference Transcripts: Approaches and Strategies,” *Library & Information Science Research* 28, no. 1 (March 1, 2006): 24–48, https://doi.org/10.1016/j.lisr.2005.11.004.
31. Julie Arnold and Neal K. Kaske, “Evaluating the Quality of a Chat Service,” *portal: Libraries and the Academy* 5, no. 2 (April 2005): 177–93, https://doi.org/10.1353/pla.2005.0017; Van Houlson, Kate McCready, and Carla Steinberg Pfahl, “A Window into Our Patron’s Needs: Analyzing Data from Chat Transcripts,” *Internet Reference Services Quarterly* 11, no. 4 (October 2006): 19–39, https://doi.org/10.1300/J136v11n04_02.
32. Peter Hernon and Charles R. McClure, “Unobtrusive Reference Testing: The 55 Percent Rule,” *Library Journal* 111, no. 7 (April 15, 1986): 37–41; Durrance, “Reference Success.”
33. Ehrenpreis and DeLooper.
34. “Ivy - Admin Portal.”
35. Sharesly Rodriguez, “Kingbot,” San Jose State University Dr. Martin Luther King, Jr. Library, 2022, https://library.sjsu.edu/kingbot; “Project Highlight: Bizzy Chat Bot,” OU Libraries, accessed August 5, 2024, https://libraries.ou.edu/content/project-highlight-bizzy-chat-bot.