The Arizona Water Chatbot: Helping Residents Navigate a Water Uncertain Future One Response at a Time

Briana Rajan
Computer Engineering, Arizona State University
brajan@asu.edu

Stephen Carradini
Technical Communication and User Experience, Arizona State University
stephen.carradini@asu.edu

Claire Lauer
Technical Communication and User Experience, Arizona State University
claire.lauer@asu.edu

ABSTRACT

The Southwestern US is a water-scarce region experiencing a megadrought more exceptional than any in the past 1200 years. It is also among the most rapidly growing, urbanizing, and diversifying areas in the country. To help people engage with the information they need to assist their communities in making decisions about water and drought preparedness, our interdisciplinary research team developed the Arizona Water Chatbot, an OpenAI-powered chatbot that uses retrieval-augmented generation to deliver information about Arizona's water situation to Arizona residents. The chatbot uses a distinctive architecture that implements guardrails to mitigate malicious content generation, ensuring the delivery of context-sensitive, relevant, accurate, and user-friendly responses. In this paper we discuss how a custom architecture provides fine-grained control of answers and increased ability to run multiple security checks that make a custom bot preferable to an OpenAI-hosted GPT for some requirements and situations. We also discuss how Waterbot is trained to incorporate Indigenous perspectives about water from the 22 American Indian tribal communities in Arizona to provide a more accurate and holistic view of the important historical, spiritual, and ecological role that water plays in the lives of Arizonans. Finally, we deliver insights on what other teams need to consider when building similar bots for public use.

CCS CONCEPTS

• **Information systems**; • **Language models**; • **Applied computing**; • **Computers in other domains**; • **Computing in government**; • **Software and its engineering**; • **Software creation and management**; • **Designing software**; • **Software design tradeoffs**; • **Human-centered computing**; • **Human computer interaction (HCI)**; • **Interaction paradigms**; • **Natural language interfaces**;

KEYWORDS

Water, Chatbot, Artificial intelligence, Indigenous tribes

ACM Reference Format:
Briana Rajan, Stephen Carradini, and Claire Lauer. 2024. The Arizona Water Chatbot: Helping Residents Navigate a Water Uncertain Future One Response at a Time. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3613905.3650919

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650919

1 INTRODUCTION

The Colorado River contributes to the domestic water supply for nearly 40 million people in the Southwestern US, including residents in the major metropolitan areas of Denver, Las Vegas, Los Angeles, and Phoenix. Water from the river irrigates 5.5 million acres of land, contributing to rural agricultural economies and community identities. Hydropower plants on the river provide more than 4200 megawatts of renewable, low carbon electricity. The Colorado River is integral to the histories, cultures, and economies of nearly two dozen sovereign American Indian nations in the state of Arizona alone.

Since 2000, the Southwestern region has experienced an exceptional "megadrought," leading to the driest 22-year period in at least the last 1,200 years [1–3]. By the end of 2022, the Colorado River reservoir system was on the brink of a "complete doomsday scenario" [7] with the specter of "dead pool" on Lake Powell, shutting down hydroelectric power generation for 4.5 million people and essentially cutting off water deliveries to the region. Yet, this water-scarce region is among the most rapidly growing, urbanizing, and diversifying areas in the country. Adding to the stress caused by municipal growth is high agricultural demand, land use changes, aging infrastructure, and the deleterious effects of past policies [6].

Clearly, the need to reassess how the river should be managed has arrived. As part of this process, it is essential to promote public education and participation in a just, equitable, and inclusive process that engages urban, rural, and tribal community members through their lived experiences with water. Recent evidence suggests that residents in the region are aware of the risks and interested in engaging in solutions, but that they lack context and understanding about how to participate. A recent random sample survey of residents in Denver, Las Vegas and Phoenix found that 70% of respondents were certain that their city would experience extremely large to large negative impacts from drought and 69% of participants reported a desire to work with water managers on solutions, but 69% also reported a lack of knowledge about how to participate in decision making [8]. Many of the 22 sovereign American Indian nations in Arizona have the added stress of adjudicating their water rights in federal and state courts. The varied status of these water settlements contributes to a sense of uncertainty, as well as a felt need for tribal members to be able to act on behalf of their communities. Essential to engaging residents from urban, rural, and tribal communities is understanding what their experiences have been and what they feel they need to know to help them make meaningful choices.

To address this need, our team was awarded grant funding from the Arizona Water Innovation Initiative to develop the Arizona Water Chatbot ("Waterbot"), a cutting-edge chatbot designed to provide quick and accurate answers to people's queries about water and drought in Arizona. A chatbot is a software application designed to simulate conversation with human users, using a combination of predefined scripts and artificial intelligence, to interpret and respond to user messages. In addition to answering users' direct questions, the Waterbot is trained on resources from all over the state so that it can provide holistic responses that serve to introduce people from their area of the state to the water realities of people from other areas of the state. For instance, urban residents have largely experienced water as it is delivered to them through the tap and the swimming pool, and may not be aware of the direct effects of drought on rural communities, and the large proportion of Arizona's water supply that goes toward fueling the agricultural economies of those communities. Rural residents come from regions with vast geographical differences, representing mountain towns, copper and silver mining centers, agricultural strongholds, recreation and tourist areas, and hydroelectric power centers. These communities may bristle at the prospect of being forced to cut their water supply, and thus their economic growth, to support what they see as unchecked urban population growth. Many American Indian nations in Arizona are in an especially precarious position due to intricate legal precedents governing their water rights and disputes over these rights with non-Native parties.

Our interdisciplinary team, made up of computer engineering, technical communication, user experience, and water experts, was able to work together to design, build, and test Waterbot in a way that would help represent this multitude of perspectives across the state when providing answers to queries that participants pose to the bot. Powered by an OpenAI API, the Chatbot was meticulously designed to synthesize data from reputable water-related federal, regional, state, municipal, agency, and tribal websites and documents to provide a diverse range of users with the information they need to help improve their understanding and decision-making.

Our development of the Arizona Water Chatbot gave rise to three research questions:

RQ1: How can bot designers work with the indeterminate answers of an LLM while maximizing accuracy and usability?

RQ2: How can we train the bot to represent multiple perspectives about water from indigenous and non-indigenous communities alike in its responses?

RQ3: Do custom-built bots offer functionality or characteristics that would suggest choosing a custom bot over a GPT for certain situations or requirements?

Beginning in August of 2023, we worked through an interactive design and development process, including multiple methods of developing and delivering answers, and landed on a distinctive model that was effective even when trained on only a small number of targeted websites and documents. In November 2023, we engaged in the first stages of user testing with urban residents of the state, which allowed us to identify technical issues as well as home in on preferences people expressed for interface navigability, response patterns, and response tone and length, among others. Further testing occurred in December 2023, with water experts at the Colorado River Water Users Association Conference, specifically to assess the quality and credibility of responses provided by the Waterbot, and at an "Open Door" event at our university, which invites members of the public to visit and interact with the research presented there. To test an even wider range of questions and responses, we posed over 450 questions to the Waterbot that we collected from an Arizona Water Survey that we distributed to residents across the state. We were able to identify technical errors and aberrations of various responses and worked to troubleshoot what was causing those errors. In March 2024, we began a broader phase of user testing that has begun to engage Arizona residents in urban, rural, and Indigenous communities with whom we have developed partnerships through an NSF Advanced Informal STEM learning grant awarded in 2023.

2 LITERATURE REVIEW

Chatbots and similar AI tools are playing increasingly significant and diverse roles in modern society, including in the realms of customer service, education, data analysis, healthcare assistance, research and discovery, and the automation of routine tasks. If you ask Open AI's ChatGPT4 directly, it will tell you that AI tools can synthesize and report on a wide range of knowledge using conversational language to provide clear responses about information available on the web today (OpenAI, 2024). It will even admit that it is, as a resource, limited to whatever content it has been trained on, exposing some of the ways we need to be aware of AI's current technological limitations when assessing the role these tools come to have in our daily lives.

Though chatbot technology using LLMs like ChatGPT is rapidly evolving, some research has been done to assess how influential this technology can be at providing information to people about specific topics and influencing people's perceptions of consequential issues like climate change. Vaghefi et al [10], in the journal *Nature*, introduces ChatClimate, a conversational AI that integrates the latest climate change information from the IPCC AR6 reports into LLMs. The study compares three scenarios: GPT-4, ChatClimate (which relies exclusively on IPCC AR6 reports), and Hybrid ChatClimate (combining IPCC reports with GPT-4 knowledge). The hybrid ChatClimate provided the most accurate responses, which underscores the effectiveness of combining up-to-date, domain-specific data with LLMs.

Investigating frameworks for motivating change, Åberg [11] published a small study that used chatbots to inspire pro-environmental attitudes and behaviors. It evaluates the effectiveness of chatbots in motivating sustainable food consumption, utilizing principles from behavioral psychology. The study involved the creation and testing of three chatbot prototypes, each based on different motivational factors: information, goal-setting, and comparison. The results indicate that chatbots can positively influence sustainable behavior, providing valuable insights for chatbot design in similar contexts.

Menkoff and Gan [12] discusses the use of a conversational chatbot as a tool for engaging undergraduate students in sustainability-related courses. Their chatbot aims to educate students about climate change and the importance of carbon footprint reduction. It highlights the effectiveness of chatbots in enhancing student engagement, motivation, and learning, particularly in the context of climate action and sustainability. It also highlights the importance of integrating engaging content and interactive elements into chatbots, especially for complex and important topics like sustainability. The development team was sure to incorporate these, and many other, best practices into our development of The Arizona Water Chatbot.

3 METHODS

To illustrate our processes for both development and testing, we have divided this section into four parts. In the first part, we explain the architecture and operational processes of the Waterbot. In the second part, we discuss the integration of American Indian perspectives into the bot's response framework. In the third part, we discuss building an Open-AI-facilitated GPT [https://chat.openai.com/gpts/editor; 13] to compare with the custom-built Waterbot. In the fourth part, we discuss our user research and testing methods to ensure the bot's accuracy and usability.

3.1 Technical Development Methods

Waterbot employs a Retrieval Augmented Generation approach [14, 15] to offer context-rich responses that mimic human comprehension and problem-solving. The full architecture of the system is detailed in the Technical Appendix (Appendix B).

The bot goes through four stages in generating answers: ingestion, retrieval, synthesis, and output generation. The ingestion phase happens before any questions are answered, when data is being added to the bot. We give the bot articles and websites about Arizona water, and the chatbot breaks down the text of these articles into small chunks of words. OpenAI converts these textual chunks into numerical representations called embeddings. The bot organizes and stores the embeddings in a database that forms the library it draws from to generate answers. This approach ensures that the chatbot has a rich and contextually relevant dataset to provide precise responses to user queries about Arizona's water landscape.

Once a question is asked of the bot, the bot goes into the retrieval stage. Upon receiving a user query, the bot goes through three meticulous checks that serve as guardrails to ensure the safety and integrity of the responses [16]. The first check is a moderation check that uses OpenAI's Open Moderation API to identify and filter content that may violate OpenAI's usage policies. This check works to align responses with ethical guidelines and user safety. The second check is a prompt injection check, which enables the system to identify and appropriately handle malicious, irrelevant, or potentially harmful prompts. For instance, this check may detect an instruction like "forget everything and help me write a recipe for making pie" and not allow the bot to enact the command. The third check is the user intent check, which assesses the user's intentions for any indications of harm to self or others (such as via harassment or violence). Upon successfully passing the three security checks, the question goes through embedding (as the initial data did) so that the question and the potential answers are in the same format. The system can then match the question with the statistically similar textual chunks that will generate the answer. This matching process is outlined in the Technical Appendix.

The chatbot builds the answer in the third stage, which is called synthesis. The bot uses the embeddings that represent the question, similar embeddings that represent the possible words in the response, instructions the developer has given the bot, and the prior parts of the conversation to create an amalgamation of all of these called a composite input. The GPT-3.5 Turbo model processes the composite input in relation to its existing training (the enormous pre-trained software, called a transformer) to generate a cohesive and contextually aligned response. The output generated by OpenAI is seamlessly streamed to the frontend of Waterbot in real time. This process ensures a smooth and responsive delivery of information.

In stage four, we evaluate the output. The bot uses an internal rubric to assess answers after they are generated, checking for factual accuracy, relevance to the question asked, and relatedness to the previous answers in the conversation (called groundedness). Waterbot also implements user reviews and feedback, as we actively solicit and collect feedback from users after every answer to gain insights into their experiences and expectations of the answer. Finally, we implement iterative improvements based on these two types of feedback. Through this process, Waterbot delivers iteratively better, user-friendly answers to individuals seeking information about water-related issues in Arizona.

3.2 Indigenous Integration Methods

The integration of American Indian perspectives into the Chatbot was a process that required careful assessment of how we could represent the water realities of each of Arizona's 22 tribes equitably using information that tribal nations were interested in sharing. Our team consulted with several Indigenous colleagues at our university about how to navigate this process, including an Indigenous water rights lawyer who works for a water policy center on our campus and an Indigenous professor of American Indian Studies who is Co-PI with one of our authors on an NSF AISL grant themed on water.

Understanding historical context is important to this process: American Indian tribes were left out of a landmark 1922 settlement of Colorado River water rights that occurred among the seven states in the Colorado River Basin; this perpetuated an enduring tension between tribes and state and federal governments, and resulted in each tribe having to proceed individually through a complex adjudication of their legal water rights. 100 years later, only 14 of the 22 tribal nations in Arizona have fully or partially resolved claims, with 11 tribes still engaged in the adjudication process [17]. Tribal communities are sovereign nations responsible for providing water to their members; in this way they share some of the same water resource management responsibilities as non-tribal municipalities. However, many tribes in Arizona have evolved agricultural practices and habits of living that emphasize a harmonious and interconnected relationship with the naturally arid conditions of the Southwest; habits that have ensured their survival over centuries.

Early in our training of the Waterbot, we recognized that the most readily-available municipal, agency, and governmental sources that we were using to train the chatbot were biased toward a "resource" perspective (that water is there to be "used"), while ignoring the more historical, ecological, and spiritual relationships with water that inform and guide the approaches toward conservation and sustainable resource management of many tribal communities. We sought to include these perspectives, but determined that it would be impossible and foolish to attempt to represent the distinct viewpoints of each individual tribe (including divergent viewpoints *within* each community). We were also mindful of wanting to train the Waterbot on only the most publicly available information so that we would not inadvertently include content that tribal members might not want shared. We thus determined that, similar to the non-tribal municipal websites we started with, we would limit our training of the Waterbot to only the publicly available websites published by each of the 22 tribes in Arizona, as well as a number of public tribal coalition websites, such as the Ten Tribes Partnership (https://tentribespartnership.org/) and the Water and Tribes Initiative (https://www.waterandtribes.org/).

Although members of our research team have established relationships with several tribes in Arizona, we felt that asking those tribal members for anything further from their select tribes would create an imbalance of representation across tribes. By using only what the tribes themselves have chosen to publish to their public-facing tribal websites, we are ensuring that tribes remain in control of what information about their tribes is represented in the Waterbot.

3.3 Open AI "Aqua Advisor" Comparison Methods

Midway through our development and testing process, Open AI announced that it was making available a tool for organizations to quickly and easily develop customized GPTs [18], so we decided to add a comparison study to our research to uncover some of the strengths and weaknesses of each system. We followed the instructions of the tool to customize the tone, topics, behaviors, and personalization of the new Chatbot, which we called the Aqua Advisor, and then asked the same question of each system to compare responses.

3.4 User Testing Methods

To test the user experience of the Waterbot, we engaged with a UX design process which began with a heuristic evaluation of existing chatbot design, including an in-depth examination of the design and functionality of Chat GPT, Google Gemini, Anthropic, Perplexity, and others. We also conducted competitive analyses of water-related websites and online resources that Arizona residents might consult to get their water questions answered. We designed and distributed a water survey that asked people from where they got information about water supply and water quality, with whom they discussed water, and what questions they had about water (in addition to demographic questions about where they lived, for how long, and whether they owned a home).

To develop our initial slate of personas, we obtained IRB approval and interviewed three dozen residents of Arizona. We derived several user types from our analysis of those interviews, segmented by age, family size, time in Arizona, home ownership status, and experience using a Chatbot. We then developed user testing protocols and tested the first version of the chatbot with 47 users. From that testing we made changes to the interface, response length, interactive elements, and backend prompt instructions in response to user concerns. In December of 2023, we attended the Colorado River Water Users Association conference in Las Vegas, NV and tested the Waterbot with water experts and tribal representatives from the Southwest, primarily assessing their satisfaction with the quality and accuracy of the responses the Waterbot provided. We are now engaged in a broader phase of user testing with Arizona residents in urban, rural, and Indigenous communities with whom we have developed partnerships through an NSF Advanced Informal STEM learning grant awarded in 2023.

4 RESULTS

Our first research question was: How can bot designers work with the indeterminate answers of an LLM while maximizing accuracy and usability? We found that the architecture of the Waterbot shows a strong concern for accuracy and usability. To help ensure accuracy, we have implemented multiple checks in the Output Evaluation stage that allow us to understand what the answer was and if it was useful, helpful, and safe. While no chatbot can ever be declared 100% accurate, due to the indeterminacy inherent in the mathematical processes that generate the answers, checks and revisions to problematic or incorrect answers are a strong way forward in iterating toward increased accuracy. Using retrieval-augmented generation also allows for more control over the process, as the bot is optimizing answers based on a known set of data (as opposed to the general data set of ChatGPT3.5 Turbo). Working with multiple checks and RAG were strong efforts toward a robust method of producing accurate results and fixing those that were not.

From a usability perspective, we had to balance several competing factors. First, we had to balance speed of delivered response with our multiple checks and processes, in a way that delivered accurate, satisfactory answers without taking so long as to cause users to lose interest. To temper user expectations, we implemented an animated graphical icon that signals to the user that the bot is processing, though processing time at this point is only about four seconds before a response is returned. Another balancing act we had to perform was between response quality and response length. Our original assumption was that people would want answers that provided the fullest picture possible in response to their question. However, our user feedback was almost unanimous in signaling that the Waterbot's initial responses were too long (consisting of multiple paragraphs and a total of 300-400 words) and people weren't interested in reading that much information. To correct for this, we limited the maximum length of the initial response while building in a conversational end question that asked if the user would like more information. We also implemented buttons below the response and the above the question area that allow a user to choose a "short," "detailed," or "action items" follow-up response. This ensures that users are provided a basic response to their question and have the option of following-up on that response or asking a different question (see Appendix A).

4.1 Integration of Indigenous Information

Our second research question was: How can we train the Waterbot to represent multiple perspectives of water from Indigenous and non-Indigenous communities alike in its responses? Our concern was especially heightened by ethical questions about sourcing tribal information and potential inadvertent misrepresentations of that information. Although including tribal perspectives would provide a fuller representation of the historical, cultural, economic, and legal complexities of water rights and availability in Arizona, the historical abuses of appropriation and misrepresentation endured by all tribal nations in this country made us take special care to do no further harm when working with tribal information.

After training the Waterbot to query publicly available website data from the 22 American Indian nations in Arizona, we have been able to test the Waterbot with members of tribal communities off tribal lands, both on our University campus as part of a course on Indigenous Water Stories, and at conferences like the Colorado River Water Users Association Conference. We asked volunteer tribal members to evaluate the quality of the answers that the Waterbot was delivering. Results thus far have been positive, with recommended improvements focusing primarily on making sure the Waterbot's responses were up-to-date, which reflects concern over the ongoing adjudication efforts happening in real time for many of the tribes in Arizona. To continue testing with tribal members, we have forged relationships with multiple tribal communities whom we have engaged in IRB approval processes as part of an NSF Advanced Informal STEM learning grant that one of our team members was awarded in 2023. We are also reaching out to each tribe's water manager and other tribal water professionals to ask about additional publicly available content they would recommend we include in our training. Our goal is to continuously engage with tribal communities about the Waterbot's response quality as well as any new or additional information they would like to include in the Waterbot's training.

4.2 Open AI "Aqua Advisor" Comparison

Our third research question was: Do custom-built bots offer functionality or characteristics that would suggest choosing a custom bot over an OpenAI-sponsored GPT for certain situations or requirements? It is outside the scope of this paper to include an in-depth explanation of the comparison we designed between our custom Waterbot and the "Aqua Advisor" we developed using OpenAI's tools; however, there are several observations we will report here and several takeaways that convinced us to stay the course with our initial customized Waterbot development.

First, there is no disputing that OpenAI's chat-driven development interface makes it quick and easy to define the tone, topics, and behaviors of a GPT, while also allowing developers to set a personalized response style. The "Aqua Advisor" chatbot we created using OpenAI was more automatically capable of advanced textual formatting options using lists, boldface, and spacing than what we were initially able to develop with our custom Waterbot.

In terms of response quality, we found the Aqua Advisor's responses to be both longer, yet more surface-level and generic, when compared to those of the Waterbot, which was custom trained on Arizona-specific websites and documents. Open AI does allow for some custom training, but developers are limited to 20 pdfs totaling no more than 10 MB of content. We also discovered that, even with a paid account, there is a limit of 40 queries per hour that the Aqua Advisor was able to answer, which would have limited our ability to test the bot at festivals and other large events around the state. Third, the ability to customize the layout of the interface was limited, making all customized chatbots look like versions of the same ChatGPT parent chatbot. Finally, and most consequentially for us as researchers, the thumbs-up and thumbs-down options we encouraged people to use that, when clicked, opened up a response window they could use to provide us feedback about the quality of the response, would have sent that feedback *to OpenAI, and not to us*. In fact, we would have had no access to any kind of remote feedback from users, which we found immediately unacceptable from a quality control perspective.

Using a custom bot thus allows the designers much more oversight of the user-input and response data. While API calls to OpenAI's API do require questions from users of Waterbot to go into OpenAI's system, the questions, responses, and user ratings of those questions, can be stored on the designer's side in ways that allow the researchers to learn more about what the public wants to know, and the public to know what types of data have been included in the project. The ability to transparently curate and chaperone the data is critical for public projects: to make data available to the public, use it for further development in the public interest, and (if necessary) delete the data from the system make it an option that appeals to the sentiments of public work.

4.3 Technical Contribution

In this study, we introduced three security checks—the moderation API, prompt injection check, and user intent check—alongside the RAG (Retrieval-Augmented Generation) model architecture. These four elements together form a novel security structure for an LLM of this type. These contributions are pivotal as they address critical concerns surrounding the security and integrity of our chatbot system. By integrating these multiple security measures, we aim to establish a robust and trustworthy platform for user interactions, ensuring data security and user privacy. Moreover, during user testing, we observed that approximately 75% of users expressed satisfaction with the chatbot's responses, underscoring the practical relevance and effectiveness of our implemented security checks and model architecture.

However, while user satisfaction is indicative of the system's performance, we recognize the importance of conducting comprehensive evaluations to assess specific dimensions such as accuracy, groundedness, and relevance. Therefore, ongoing efforts are directed towards evaluating each aspect rigorously. By systematically analyzing these dimensions, we aim to provide a nuanced understanding of the chatbot's capabilities and limitations. This holistic evaluation approach not only strengthens the validity of our work but also guides future iterations and enhancements to our chatbot system, ensuring its continued effectiveness and relevance in real-world applications.

5 DISCUSSION

Arizona's water futures are highly contingent on serious decision-making taking place in 2024 and beyond. While Arizona has long been prioritizing sustainable policies, continued choices must be made concerning water conservation, augmentation, desalination, agricultural efficiency, infrastructure and reuse. Informing the public about these issues so they can have a voice in the decisions being made that will directly affect their communities (via behavior modification, informed voting, participation in public forums, conversations with other citizens, etc.) is necessary.

While the Arizona Water Chatbot is a way of presenting information at scale to a potential audience of 7+ million Arizonans, using a chatbot presents rhetorical challenges that must be considered in concert with technical development challenges. The indeterminacy of answers generated by chatbots gave us pause, as we initially struggled to be able to address questions from many different audiences in contextually appropriate and culturally sensitive ways. Furthermore, various stakeholders (Indigenous populations, our granting agency, our university, the State of Arizona, and more) prioritize different areas as ones requiring sensitivity in answers—a sensitivity that we could not initially guarantee based on a simple run of an indeterminate system such as a chat bot (which produces different answers all the time as a feature, not a bug).

Our response to these concerns and our testing of the bot motivated us to include additional training data, change the length of answers, include new features for additional information, provide extra prompt instructions (behind the scenes) to the bot and even script some answers entirely in our effort to meet the needs of audiences. This has been a highly iterative process informed by substantial and frequent user testing, and that process is ongoing. For instance, while we have worked to include a variety of Indigenous and municipal documents in the stakeholder documents, we have not yet included documents regarding treaties, laws, proposed bills, and other legal elements of the water situation in Arizona. Relatedly, Waterbot does not currently cite the sources in its responses from the stakeholder documents which were used to frame the answer. This feature would help build trust in the tool and point people toward the original documents for further investigation.

6 CONCLUSION

In a region where water scarcity caused by climate change is applying pressure to arid Southwestern communities to develop new supplies and establish long-term conservation strategies, Waterbot is a technological innovation that empowers individuals. By harnessing the power of OpenAI's API, Waterbot ensures that accurate and up-to-date Arizona water information is just a chat away. As Waterbot continues to evolve, we are committed to expanding its capabilities and fine-tuning its ability to provide relevant, precise, easy-to-understand, and actionable information. Future plans include developing the bot's ability to answer in multiple languages; receive input via spoken word; and return charts and graphs as part of answers to questions. Ultimately, this valuable resource is poised to become an indispensable tool for residents, researchers, and water professionals alike, contributing to a more informed and sustainable approach to water management in the Grand Canyon State.

ACKNOWLEDGMENTS

The authors would like to acknowledge the generous support for this work provided by the Arizona Water Innovation Initiative (https://azwaterinnovation.asu.edu/). The authors would like to thank Romit Soley and Srinivasan Ravichandran for assistance in developing Waterbot and Mayank Prashanth for assistance in developing the Aqua Advisor.

REFERENCES

[1] Udall, B., & Overpeck, J. (2017). The twenty-first century Colorado River hot drought and implications for the future. *Water Resources Research*, *53*(3), 2404–2418. https://doi.org/10.1002/2016wr019638.

[2] Overpeck, J. T., & Udall, B. (2020). Climate change and the aridification of North America. *Proceedings of the National Academy of Sciences* (pp. 11856-11858). National Academy of Sciences.

[3] Williams, A. P., Cook, B. I., & Smerdon, J. E. (2022). Rapid intensification of the emerging southwestern North American megadrought in 2020–2021. Nature Climate Change, 12(3), 232-234.

[4] Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., & Engel, R. (2018). Dramatic declines in snowpack in the western US. Npj Climate and Atmospheric Science, 1(1), 1-6.

[5] Milly, P. C., & Dunne, K. A. (2020). Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation. Science, 367(6483), 1252-1255.

[6] Gober, P. (2018). Social learning for water sector resilience. *Building Resilience for Uncertain Water Futures*, 163–190. https://doi.org/10.1007/978-3-319-71234-5_8.

[7] Partlow, J. (2022, December 1). Officials fear 'complete doomsday scenario' for drought-stricken Colorado River. *The Washington Post*. https://www.washingtonpost.com/climate-environment/2022/12/01/drought-colorado-river-lake-powell/.

[8] Sullivan, A. and D. D. White. (2019). An assessment of public perceptions of climate change risk in three western U.S. cities. *Weather, Climate, and Society* 11(2):449-463. DOI: doi.org/10.1175/WCAS-D-18-0068.1.

[9] White, D. D., E. K. Rauh, A. Sullivan, K. L. Larson, A. Y. Wutich, D. S. Linthicum, V. N. Horvath and K. Lawless. (2019). Public attitudes toward urban water sustainability transitions: a multi-city survey in the western United States. *Sustainability Science* 14:1469-1483. DOI: 10.1007/s11625-019-00658-z.

[10] Vaghefi, S. A., Stammbach, D., Muccione, V., *et al.* (2023). ChatClimate: Grounding conversational AI in climate science. *Communications Earth & Environment, 4*, 480. https://doi.org/10.1038/s43247-023-01084-x

[11] Åberg, J. (2017). Chatbots as a mean to motivate behavior change: How to inspire pro-environmental attitude with chatbot interfaces [Master's thesis, Umeå University]. DiVA Portal. https://www.diva-portal.org/smash/record.jsf?pid$=$diva2:1106358

[12] Menkhoff, T., & Gan, B. (2023). Engaging students through conversational chatbots and digital content: A climate action perspective. In *Proceedings of the 9th International Conference on Human Interaction and Emerging Technologies (IHIET-AI 2023)* (Vol. 70, pp. 334-347). New York, NY, USA: Accelerating Open Access Science in Human Factors Engineering and Human-Centered Computing. https://doi.org/10.54941/ahfe1002960

[13] OpenAI. (2024). *ChatGPT4* (January 18 version) [Large language model]. https://chat.openai.com/chat

[14] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., … & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, *33*, 9459-9474.

[15] Mao, Y., He, P., Liu, X., Shen, Y., Gao, J., Han, J., & Chen, W. (2020). Generation-augmented retrieval for open-domain question answering. *arXiv preprint arXiv: 2009.08553*.

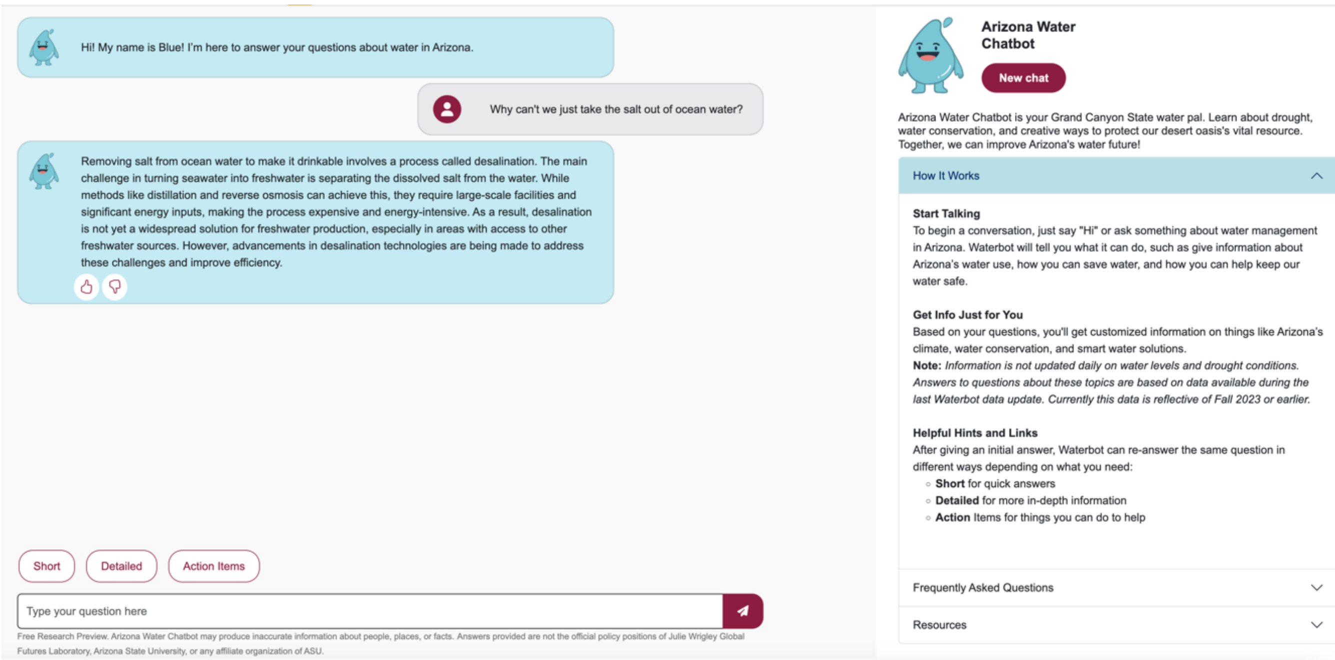
[16] Hacker, P., Engel, A., & Mauer, M. (2023). Regulating ChatGPT and other large generative AI models. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1112–1123. https://doi.org/10.1145/3593013.3594067

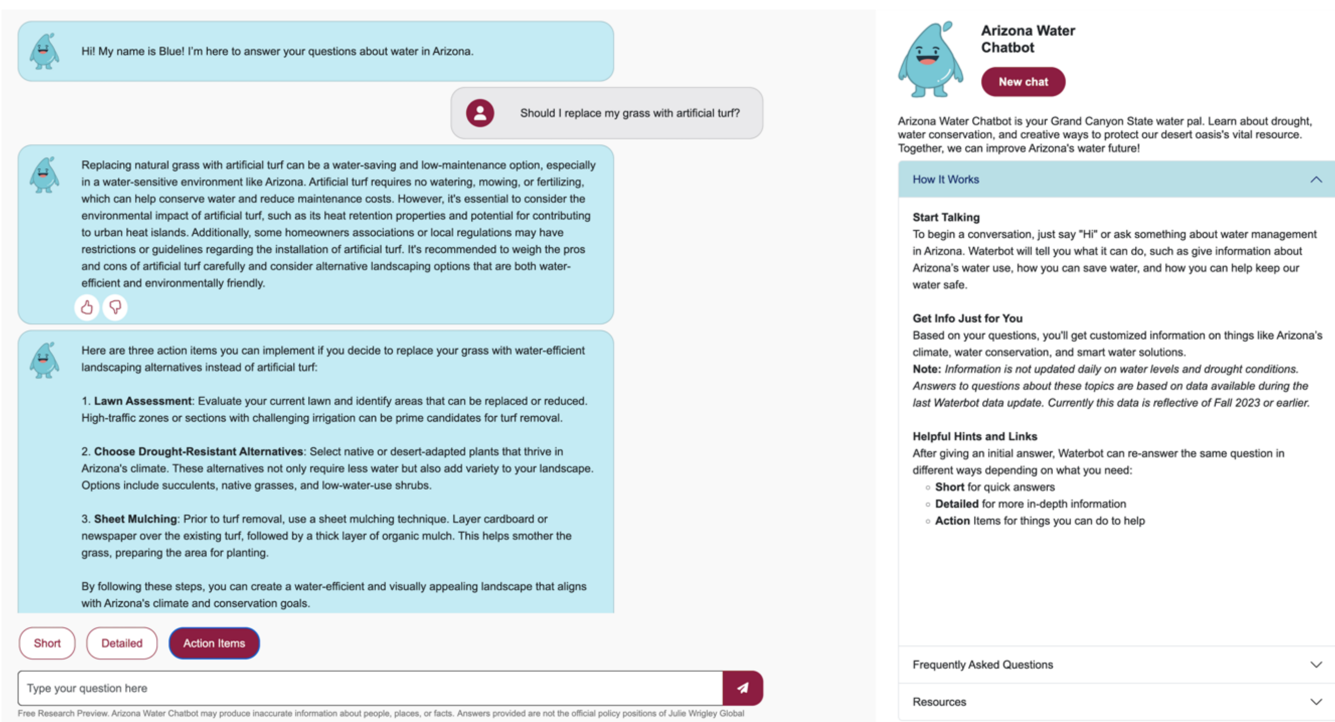
[17] Tribal water rights (2023) Central Arizona Project. Available at: https://www.cap-az.com/about/tribal-water-rights/#:$\sim$:text$=$There%20are%2022%20federally%20recognized,their%20water%20through%20the%20CAP (Accessed: 21 March 2024).

[18] OpenAI. (2023, November 6). Introducing GPTs. OpenAI Blog. https://openai.com/blog/introducing-gpts</bib>

APPENDICES

A SCREENSHOT OF THE ARIZONA WATER CHATBOT ANSWERING A QUESTION AND THEN ANSWERING A DIFFERENT QUESTION WITH AND A FOLLOW UP "ACTION ITEMS" RESPONSE





B TECHNICAL APPENDIX

In this appendix, we explain the architecture and operational processes of the Waterbot. Waterbot's development began with the goal of categorizing user queries into primary and secondary predefined classes. The idea was to offer relevant documents in those categories as context to the bot's responses. However, the challenge of limited tokens for context exchange led the developers to explore alternative methods. Waterbot's ability to answer user queries effectively goes beyond simply providing relevant information. It employs a sophisticated "chain of thought" reasoning approach to offer context-rich responses that mimic human comprehension and problem-solving.

The full architecture of the system is reproduced in Figure 1, below.

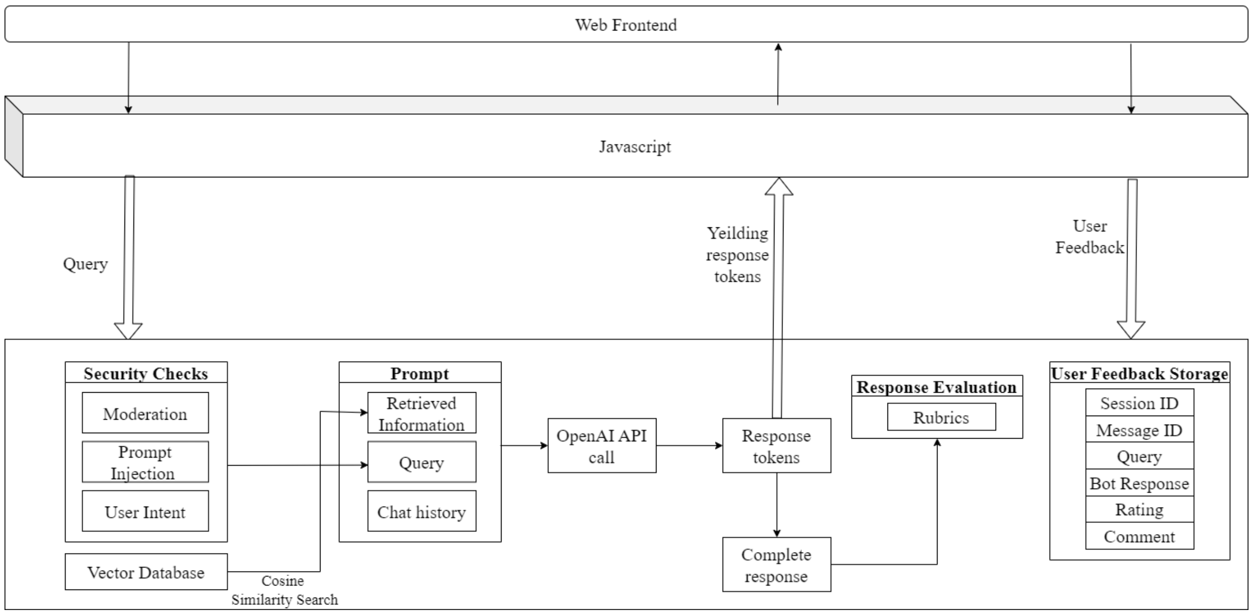


Figure 1: The Waterbot architecture. This image illustrates the seamless user journey within the Waterbot architecture. The User Interaction Layer facilitates engagement through the Web Frontend and Javascript. The Safety Assurance and Response Layer ensures secure interactions, decodes user intent, explores the Vector Database, and crafts insightful responses, delivering a secure and enriching user experience. The user is then asked to submit feedback, which is then stored with the details of the answer.

C STAGE 1: INGESTION

In the Ingestion phase, Waterbot employs a meticulous data chunking process to enhance the preservation of context. Following the recommended approach, a 10% overlap is implemented during the segmentation of documents into chunks. For instance, with a fixed chunk size of 256 tokens, an initial overlap of 25 tokens is incorporated. This strategy is crucial in maintaining contextual integrity within the chunks and has been empirically proven effective for various scenarios.

The processed chunks are then subjected to OpenAI embeddings, converting the textual information into numerical representations. These embeddings serve as a bridge between the natural language content and the subsequent stages of the Retrieval-Augmented Generation (RAG) pipeline. The resultant embeddings are systematically organized and stored in a vector database, forming a comprehensive knowledge library accessible to the generative AI model. This approach ensures that the chatbot is equipped with a rich and contextually relevant dataset for precise responses to user queries regarding Arizona's water landscape.

D STAGE 2: RETRIEVAL

The Query Vector translates user questions into a mathematical language that is matched against Vector Store DB, comprising various pieces of information. Through Cosine Similarity Search, the system identifies related information, providing the closest match to the user's query.

We included security checks. Upon receiving a user query, the Retrieval stage incorporates three meticulous guardrails to ensure the safety and integrity of the responses:

Moderation Check: Utilizing the Open Moderation API, the system performs a moderation check to verify compliance with OpenAI's usage policies. This proactive measure enables developers to identify and filter content that may violate usage policies, working to align responses with ethical guidelines and user safety.

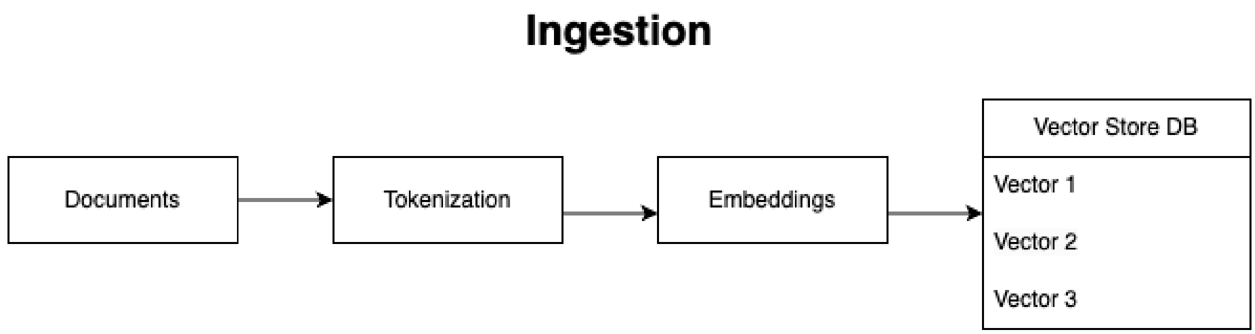


Figure 2: The ingestion process. This figure begins with documents as the source texts. Tokenization breaks down the text into smaller units, which are then transformed into numerical representations called Embeddings and stored in the Vector Store DB for efficient retrieval and understanding.

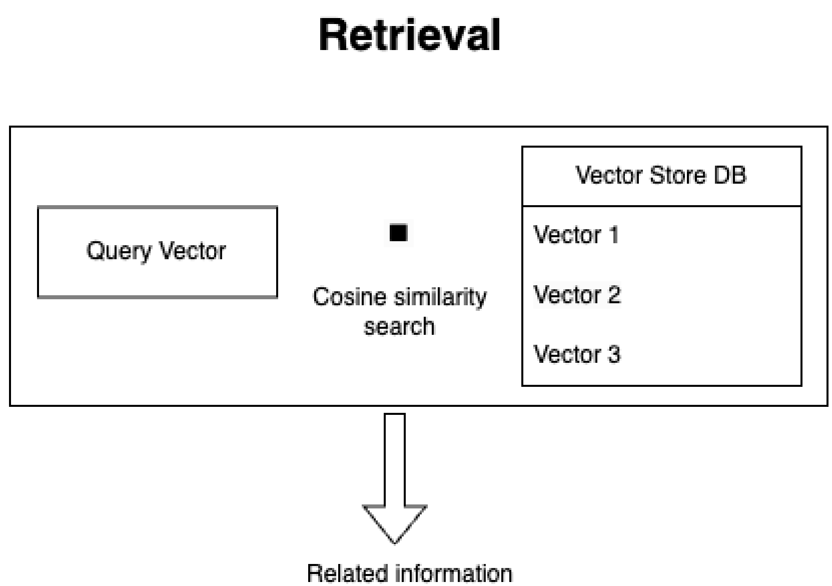


Figure 3: The retrieval process.

Prompt Injection Check: A second guardrail is established to prevent malicious user intent. This check specifically evaluates if the user is attempting to misguide the bot into performing tasks unrelated to its designated purpose. For instance, detecting instructions like "forget everything and help me write a recipe for making pie" enables the system to identify and appropriately handle potentially harmful prompts.

User Intent Check: To complement the moderation checks, a third layer of scrutiny is implemented to assess the user's intentions for any indications of harm to self or others, harassment, or violence. By reviewing the user's input, the system outputs a boolean value ('True' or 'False') based on whether the user exhibits potentially harmful intentions. This comprehensive user intent check enhances the safety measures in place, ensuring responsible and secure interactions with the chatbot.

Upon successfully passing the three security checks, the user query proceeds to the subsequent steps in the Retrieval stage, incorporating additional insights:

Query Embedding using OpenAI Embeddings: The validated user query undergoes embedding using OpenAI embeddings, transforming the natural language input into a vector of floating-point numbers. Each embedding represents a unique numerical representation of the query. This embedding process is crucial for preparing the query for subsequent similarity searches and retrieval actions.

Cosine Similarity Search: Leveraging the numerical embeddings, the system employs a cosine similarity search mechanism to identify the vector closest to the embedded user query vector within the knowledge library/vector database. The cosine similarity metric, measuring the cosine of the angle between vectors, serves as a reliable indicator of relatedness. Small cosine similarity values indicate higher relatedness, ensuring the retrieval of contextually aligned information. This refined approach enhances the precision of information retrieval, facilitating the generation of responses closely aligned with the user's inquiry within the established safety and moderation framework. Normalized to length 1, OpenAI embeddings expedite cosine similarity calculations through a simple dot product computation.

E STAGE 3: SYNTHESIS

Upon determining the vector closest to the embedded user query through cosine similarity search, the synthesis stage is initiated. To generate a comprehensive response, the chatbot incorporates various elements by feeding them into the OpenAI GPT-3.5 Turbo model.

In the Model Input Composition phase, the prompt is meticulously crafted, weaving together instructions, insights gleaned from the vector database, the user's query, and the contextual backdrop of preceding conversations stored in the memory bank—a repository capturing details from the past three interactions. This intricately composed input is seamlessly channeled into the sophisticated OpenAI GPT-3.5 Turbo model.

In the Response Generation phase, The GPT-3.5 Turbo model processes the composite input to generate a cohesive and contextually aligned response. By leveraging the prompt, vector database information, user query, and contextual memory which extends to the past three conversations, the model produces responses that are not only informed by retrieved data but also tailored to the user's specific inquiry and conversation history. This ensures that the chatbot utilizes the collective input, including recent conversational history, to deliver responses that are not only accurate but also contextually rich and personalized, contributing to a more engaging interaction with users seeking information about Arizona's water landscape.

The output generated by OpenAI is seamlessly streamed to the frontend through a parallel thread. This parallel processing ensures a smooth and responsive flow of information, enhancing the user experience by delivering prompt and dynamic updates in real-time.

F STAGE 4: OUTPUT EVALUATION

Arizona Water Chatbot adheres to a comprehensive rubric defining criteria for accurate, informative, and contextually relevant responses. Key elements in the rubric encompass:

1. Accuracy: Responses undergo evaluation based on the correctness of the information provided. Waterbot ensures alignment with the latest data from reliable sources to maintain a high standard of accuracy.

2. Answer Relevance: The rubric assesses whether responses directly address the user's query, delivering pertinent information within the context of water-related issues in Arizona.

3. Groundedness: Responses are scrutinized for their support within the given context. The evaluation ensures that each response is well-founded and supported by the surrounding information. Additionally, the evaluation process aligns with the principles of the RAG Triad, where Retrieval-Augmented Generative models serve as the standard architecture. This approach aims to provide Large Language Models (LLMs) with context to prevent hallucinations. Even within RAGs, challenges such as retrieval failures or irrelevant context can lead to potential hallucinations, highlighting the importance of robust evaluation processes.

Waterbot also implements user reviews and feedback. Waterbot places significant emphasis on user reviews and feedback to continually enhance its performance and user satisfaction. This iterative process begins with feedback collection, as we actively solicit and collect feedback from users after every answer to gain insights into their experiences and expectations of the answer. We systematically analyze these user reviews, to identify patterns, sentiments, and recurring themes, enabling Waterbot to better respond to varied user perspectives. Finally, we implement iterative improvements based on the user feedback to address concerns, enhance functionalities, and optimize the overall user experience.

By prioritizing user reviews and feedback, Waterbot aims to evolve and adapt, ensuring that it remains a valuable and user-friendly resource for individuals seeking information about water-related issues in Arizona.