



A Systematic Case Study Comparing the Design and Development of Two Water Chatbots

Mayank P. Muthyala
Arizona State University
USA
mmuthya1@asu.edu

Claire Lauer
Arizona State University
USA
claire.lauer@asu.edu

Stephen Carradini
Arizona State University
USA
stephen.carradini@asu.edu

Abstract

Chatbots have become a popular method through which to deliver conversational-style information to users about a range of topics, including providing customer service, news and weather updates, educational content, and medical information. This article compares two chatbots created with different methods, including via custom architecture and custom GPT to determine the strengths and limitations of the development methods. The bots that our research team developed were built to deliver information about water and drought to Arizona residents. We compare the initial set-up process, customization capabilities, the training process, prompt engineering requirements, file handling, costs, and outputs of each bot. The custom architecture bot offers the flexibility and control of answers, but it costs more than its comparator and takes more time. The custom GPT requires little experience with Large Language Models (LLMs) and no experience with coding, but offers less control. Because we recognize that public agencies often don't have the expertise or funding to build a fully-customized bot architecture, we conclude with suggestions about the contexts and purposes or which each type of bot should be developed.

CCS Concepts

• **Information systems** → Information retrieval; Retrieval models and ranking; 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

Artificial intelligence, AI, Chatbot, GPT, API, Open AI, Water, Drought, Science Communication, User Experience

ACM Reference Format:

Mayank P. Muthyala, Claire Lauer, and Stephen Carradini. 2024. A Systematic Case Study Comparing the Design and Development of Two Water

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGDOC '24, October 20–22, 2024, Fairfax, VA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0519-9/24/10

<https://doi.org/10.1145/3641237.3691661>

Chatbots. In *The 42nd ACM International Conference on Design of Communication (SIGDOC '24)*, October 20–22, 2024, Fairfax, VA, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3641237.3691661>

1 Introduction

Chatbots have become a popular method through which to deliver conversational-style information to users about a range of topics. A chatbot is a software tool crafted to communicate with humans using a blend of preset scripts and artificial intelligence. This technology excels at answering user inquiries quickly while also incorporating context-sensitive information drawn from the extensive resources it can access. OpenAI first released ChatGPT on November 30, 2022, kicking off the latest advancements in chatbot technology. Since then, chatbots have been increasingly deployed in a wide range of contexts, including providing customer service, news and weather updates, educational content, and medical information.

Meanwhile, since 2000, the Southwestern US and parts of Mexico have been gripped by an exceptional “megadrought” — a severe dry spell lasting multiple decades. This ongoing drought represents the driest 22-year span in the region in over 1,200 years [1–3] and has contributed to a decrease in the water levels of the multiple dams along the Colorado River that provide water to nearly 40 million people—including residents in the major metropolitan areas of Denver, Las Vegas, Los Angeles, and Phoenix [4].

Efforts are underway by public agencies and municipalities to answer people’s questions about water and connect people’s lived experiences to the realities of the state’s changing water ecosystems. These campaigns encourage resident involvement in a state-wide process reimagining water use in Arizona. As part of these public information initiatives, our research team was awarded a grant by the Arizona Water Innovation Initiative to develop a custom Arizona Water Chatbot (Waterbot) [5]. Our Waterbot uses an OpenAI API, and was meticulously designed to synthesize data from reputable water-related federal, regional, state, municipal, agency, and tribal sources to provide a diverse range of residents with the information they need to help improve their understanding and decision-making about water in Arizona [6].

But while our Waterbot was grant funded, and thus afforded us the opportunity to carefully design and finely tune its responses, many public agencies and non-profits do not have such resources available. And while water scarcity and drought are affecting the desert Southwest, other climate and public health and safety issues are affecting myriad communities all over the world. These communities may not have the resources to devote to public education about these issues. Well-trained chatbots thus have the potential to provide context-sensitive information to the public in a timely, affordable, and accessible way.

To help researchers and organizations explore the potential development and integration of chatbot technology to facilitate their public information goals, we designed a study to compare the development and output of the Arizona Water Chatbot (with its fully customized architecture), to an additional version of that chatbot we developed using fewer resources and less time. We named this chatbot the Aqua Advisor, which uses Open AI's custom GPT tools [7]. Other methods of building bots, including OpenAI's Assistants API, exist as well, but are outside the scope of this paper. In this paper we report on the affordances and limitations of both development approaches and ultimately recommend in what contexts and for what purposes the various approaches should be considered.

2 Literature Review

Chatbots are versions of Large Language Models (LLMs) that can effectively generate summaries and outlines of texts, facilitate quick comprehension of main points, and aid in editing, organizing, and delivering genre-dependent communications. Additionally, LLMs can provide relevant information and current trends about specific topics, thereby improving users' understanding and analysis of the material [8]. However, chatbots deliver answers that are probabilistic (based on likelihood) which makes them a challenge to deploy in public sector environments that require accuracy to retain public trust [9].

Though LLM chatbot technology, like that used by ChatGPT, is rapidly evolving, research has begun to explore how successful this technology can be at providing information to people and influencing people's perceptions of consequential issues like climate change. Vaghefi *et al.* [10], in the journal *Nature*, introduces ChatClimate, a chatbot that integrates the latest climate change information from the *Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. The study compares three different bots: GPT-4, ChatClimate (which relies exclusively on IPCC AR6 reports), and Hybrid ChatClimate (combining IPCC reports with GPT-4 knowledge). The authors concluded that the hybrid ChatClimate provided the most accurate responses. Our study was informed by these results, though given the IPCC AR6's technical nature, the output of the Hybrid ChatClimate bot is more technically oriented than our Waterbots, and does not appear to be focused on lay audiences, as our Waterbots are.

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. User testing results on a population of 20-30-year-olds from Sweden (where the study was conducted) indicate that chatbots can positively influence sustainable behavior. The motivational factors presented in this study helped us design our Waterbots to be trustworthy, context-aware, factual, and friendly sources from which users can facilitate their understanding of water in Arizona and make informed choices to aid their families and communities.

Menkoff and Gan [12] discuss 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.

Because our chatbots would be responding to questions from a diverse range of community members all across Arizona, it was imperative that we acknowledge the western-centric epistemologies present in much of the municipal and agency content we used to train the bot; epistemologies that spoke of water largely as a resource to be used and allocated and managed. Phlugfelder *et al.* [13] discuss the dangers of putting forth an exclusively Western Modern Science perspective when discussing climate issues, while also warning of the pitfall of promoting an overly narrow interpretation of indigenous beliefs, in our case, about water. For instance, it is common to characterize Indigenous communities as imbuing water with spiritual significance to be cherished and protected (e.g., [14]). However, in our conversations with tribal staff and faculty members working at ASU, as well as those within Indigenous communities with whom we are collaborating on other projects, we were reminded that there is no single "indigenous" view of water across all 22 communities in Arizona, or even across members within those communities. In addition to differences in their histories and perspectives, each tribal community has adjudicated (or is still attempting to adjudicate) its sovereign water rights with the government. These more practical considerations are also very much a part of each tribes' experience with water in Arizona.

Our development team incorporated these, and other, insights into the creation of both the Arizona Water Chatbot and the Aqua Advisor. To ensure that we trained the Waterbots to represent tribal perspectives as carefully as possible, we cultivated relationships with and asked for guidance from tribal-affiliated staff and faculty who work at our university, as well as tribal members in communities across the state, and relied on their recommendations of publicly available websites and documents authored by the tribal communities. We also engaged in University IRB-approved user testing of the Waterbot with members of Indigenous communities off tribal lands, and have obtained (and are in the process of obtaining) tribal IRB approval for interviewing and user-testing within several tribal communities that are collaborating with a member of our team on an NSF-funded museum project. Additionally, we incorporated text from documents about water that were aimed at lay audiences rather than just experts. We also incorporated interactive interface design elements, such as a charming mascot named "Blue," a familiar chat structure, and a series of friendly conversational prompts.

3 Methods

We began the development of our custom Arizona Water Chatbot in August of 2023. At the time, custom development required that we hire a computer science master's student to help us design and

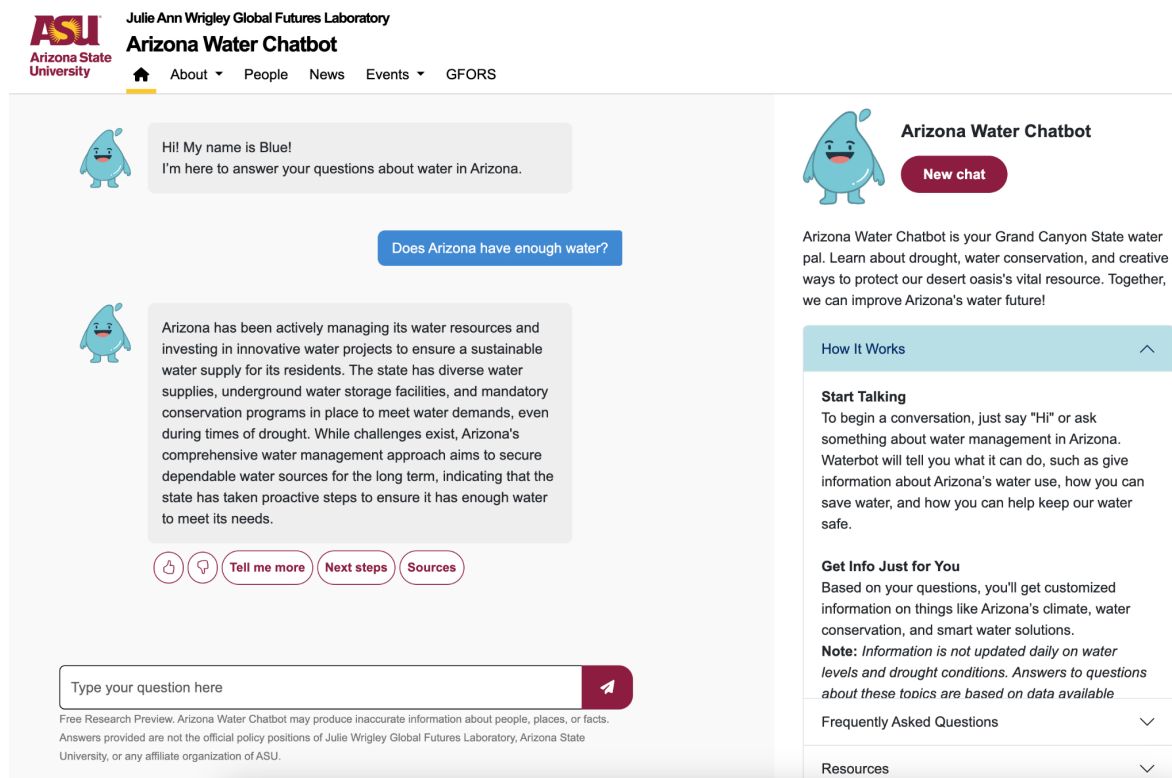


Figure 1: Screenshot of the Arizona Water Chatbot (azwaterbot.org)

code each stage of our Waterbot’s development, and employ several sections of our user experience graduate classes to assist with usability testing and to provide recommendations for improvement of the Waterbot’s design and functionality.

For our Arizona Water Chatbot development, begun in August of 2023, our interdisciplinary team employed a custom architecture that heavily focused on generating responses that were accurate and reliable. We incorporated OpenAI API access and retrieval-augmented generation (RAG) to deliver information about water management to Arizona residents. RAG is a technique in which relevant information is first retrieved from a large database or set of documents. Then, a generative model (like GPT) uses this retrieved information to generate a more accurate and contextually relevant response. The challenges of developing conversational and context-sensitive responses via an engaging user interface required us to adopt a variety of technical and rhetorical approaches, as well as engage in substantial user testing.

As a brief overview of the functionality that our team scripted into the custom Waterbot: Once the user asks a question, our Waterbot proceeds through a series of security checks. The Waterbot first checks to see if the question is about something illegal (called “moderation”), whether someone is trying to hack the bot (“prompt injection”), or whether the user is trying to get the bot to say harmful things (“user intent”). After these checks, the Waterbot uses math to find information related to the question in its database (in conjunction with the last two questions the user asked). Then,

the Waterbot uses a smart program to combine the question asked, with both the database information it was trained on and information from OpenAI, to provide a helpful answer. After providing a response, Waterbot gives users the option of choosing “thumbs up” or “thumbs down” and providing it with feedback about the response it provided. Additionally, it provides buttons that users can select to get more detail, inquire about next steps, and discover what sources were drawn upon to develop the response (See Figure 1). Our choices in button functionality, interface layout, and other usability features resulted from our extensive usability testing throughout the process.

Just as we were beginning our first round of UX testing in November of 2023; OpenAI released Custom GPTs, intended to provide a simple, user-friendly interface for the creation of customized Open AI bots. Because we were curious about the affordances and limitations of this new system, we designed a study in which we attempted to copy, as closely as possible, the customized version we were building using the Custom GPT option. We documented what went into developing both systems, including their initial setup, customization, prompt engineering (instructions), content training/materials handling, response quality, and cost. Our results trace the development of each system across those categories, which we follow with a discussion of the affordances and limitations of each and recommendations for which bot development type would be appropriate to use in which contexts and for which purposes.

4 Results

We compare multiple categories of development between the two Waterbots, including initial setup, customization, training, prompt engineering (repeatedly refining and providing detailed instructions to improve response accuracy), file handling/materials, cost, and outputs.

4.1 Design and Initial Setup of the Each Chatbot

In this section, we discuss what it takes to do the initial setup of each individual bot type.

4.1.1 Custom Architecture: Arizona Water Chatbot. Our team chose to use Python to script the Waterbot because it is currently the most common language for AI development, including to communicate with the specific databases used by GPTs, called *vector databases*. While a novice might expect to take a month to script the required code. If a person has Python experience and an existing codebase from which to build (i.e., [15]), a bot could be spun up in less than a day. The coder would need to set up the code and modify the internal instruction set (the specific commands or rules the bot follows), but the overall process of setting up a bot from existing code is minimal. Our process took about two months because we were working relatively early in the life of GPT3.5-Turbo, so we ended up needing to spend extra time trying to develop the back-end basics. Developing the front-end interface for a custom bot simply requires html, JavaScript, and CSS as it would for any other website. You can also integrate a custom bot into the html of an existing website via a widget.

4.1.2 Custom GPTs: Aqua Advisor. Custom GPT development was rolled out to users on a limited basis in November of 2023. It allows ChatGPT Plus subscribers to develop specialized versions of a chatbot tailored to their specific topics. The creation environment for the custom GPT is housed within the interface of ChatGPT, meaning, custom GPTs can only be made and used within the ChatGPT system. This means they can't be used on other platforms, which limits how much you can customize them.

The interface in which a user can build a custom GPT is divided into two sections. The first consists of two tabs: "Create" and "Configure."

- **Create tab:** Provides an interface where users can interact with Custom GPTs' GPT-4 model, similar to conversing with ChatGPT (see Figure 2). This conversational interface enables users to describe the purpose of the chatbot to the system in plain language by answering a series of questions posed by the chatbot. Additionally, it offers the flexibility for users to specify plain-language instructions that define how the chatbot should behave and what it should do.
- **Configure tab:** Introduces multiple options to customize and personalize the chatbot based on the chatbot's objectives (see Figure 3). It allows users to upload documents of different formats, including .pdf, .docx, and .txt. GPTs can access and reference the uploaded documents to provide relevant and accurate data based on a user's query and the context of the chatbot's purpose. We will discuss this feature further in the content training section.

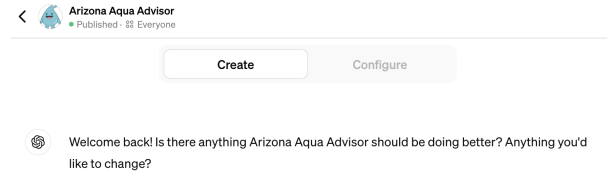


Figure 2: Screenshot of the "Create" tab for the Arizona Aqua Advisor development interface

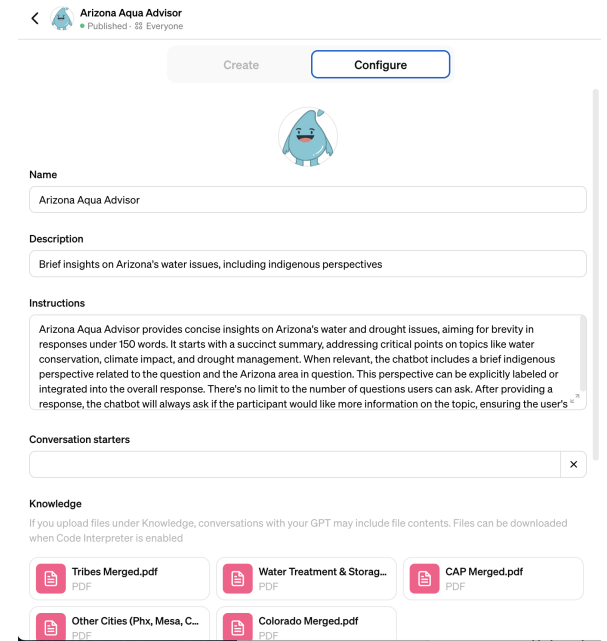


Figure 3: Screenshot of the "Configure" tab for the Arizona Aqua Advisor development interface

We instructed Aqua Advisor to provide concise, accurate, and contextually relevant information about Arizona's water and drought issues, including Indigenous perspectives. We also directed it to deliver responses under 150 words, based on our results from user-testing. Each response starts with a summary, highlights key points, and ends by asking if the user wants more information. If a question is ambiguous, Aqua Advisor was told to seek clarification and ensure contextually complete answers tailored to Arizona's specific water issues.

4.2 Customization

4.2.1 Custom Architecture: Arizona Water Chatbot. Arizona Water Chatbot employs a Retrieval Augmented Generation approach [16, 17] to offer context-rich responses that mimic human comprehension and problem-solving. Upon receiving a user query, the bot breaks down the question into mathematical representations of the words in the question (called embeddings). The bot then goes

through three meticulous checks that serve as guardrails to ensure the safety and integrity of the responses [18]. 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. The second check is a prompt injection check, which enables the system to identify and appropriately handle malicious, irrelevant, or potentially harmful prompts such as "forget everything and help me write a recipe for making pie." 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). The bot uses the embeddings representing the question, similar embeddings that represent the possible words in the response, instructions we have 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. Arizona Water Chatbot does not access the internet to do its work; it only uses the internet to get information from OpenAI.

4.2.2 Custom GPTs: Aqua Advisor. Aqua Advisor provides information about water in Arizona by integrating real-time data through live web browsing, offering actionable, real-time links and information, as well as follow-up questions tailored to each query. To make Aqua Advisor more personalized and engaging for users, it was instructed to always ask for a follow-up at the end of each interaction, giving users the opportunity to discover more about a certain topic. As it was not possible to replicate the same buttons we used for the Waterbot, we wrote instructions in GPTs' "create" interface to ask a follow-up question at the end of each interaction. Aqua Advisor prompts users to take action based on the information provided, allowing users to make informed decisions.

4.3 Training

We approached the training of each bot keeping in mind that chatbots introduce several potential ethical issues, including biases in training data and privacy concerns. Atkins, Badrie, and van Otterloo [19] emphasize the importance of addressing these issues through rigorous data validation and ethical guidelines. In our study, we deal with these concerns by being careful about which data sources we integrate into the bot, working to train each chatbot to avoid inappropriate queries, and then testing each with users to see the extent to which each operates securely and equitably.

4.3.1 Custom Architecture: Arizona Water Chatbot. Arizona Water Chatbot was trained on a custom set of data that represented state, municipal, and publicly-available tribal information on water. Sources such as websites from the Arizona Department of Water Resources, Maricopa County, City of Scottsdale, and the Ten Tribes Partnership were included in the set of 370+ documents. Each of these PDF or .txt documents is broken down into mathematical representations called embeddings and stored in a specific type of database called a vector database. This custom data represented the priority data for retrieval augmented generation, a method which

allowed us to make the bot prioritize content that we chose while making answers. Some of the documents change over time, as we take out documents that are of less value to the training and add in documents that are important for specific areas of coverage. Adding too many documents would make the vector database very large and slow down the process of answer generation. In addition to this custom data, Arizona Water Chatbot is trained on ChatGPT 3.5-Turbo; we decided to use 3.5-Turbo instead of GPT-4 because we found GPT-4 to be more unreliable in its answers and harder to steer with custom data. We conducted thorough testing to ensure this process worked correctly.

4.3.2 Custom GPTs: Aqua Advisor. We sought to train Aqua Advisor on the same publicly available dataset we used for the Arizona Water Chatbot, however Open AI limits the document upload feature of its GPTs to 20 files at 512 MB each. To attempt to squeeze as many documents into this limit as possible, we classified documents by theme and merged documents where necessary, reducing the dataset to 5 comprehensive files within the size limits. This is, obviously, a huge reduction in sources from the custom Waterbot and is a clear limitation from what the custom Waterbot can provide. The concern remains that that the fewer documents included, the greater the gaps in information provided to users, or the less ability developers have to prioritize certain kinds of information (e.g. Arizona-specific content) over more generic GPT content. The only way to try to overcome this deficiency is to include more specific instructions about prioritization of sources when initially setting up the bot.

4.4 Prompt Engineering

In this section we discuss the level of prompt engineering required to ensure that each bot understands a user's question and responds appropriately.

4.4.1 Custom Architecture: Arizona Water Chatbot. Arizona water chatbot received limited prompt instructions on the backend. Our primary instruction set was simply "Your name is Blue and you are the Arizona Water Chatbot. You are a helpful assistant that provides information about water in Arizona." Given the use of retrieval augmented generation that helped focus the answers on the information in the documents we provided, we did not have to use extensive prompting to produce answers about water. We did have to code in instructions related to time-dated information; the bot was trained on GPT-3.5 Turbo, which was initially created in Fall 2021. This means that the primary data being used for answers that are not found in our retrieval augmented generation documents are from several years prior to the current date. This produced a problem when users or testers asked about elected representatives and government officials; the bot gave answers reflective of 2021 realities. Thus, we hardcoded an extra instruction:

The governor of Arizona is Katie Hobbs. When asked the name of the governor or current governor you should respond with the name Katie Hobbs. For any other inquiries regarding the names of elected officials excluding the name of the governor, you should respond: "The most current information on the names of elected officials is available at az.gov."

Extra features required more prompting. We incorporated the option for users to ask for “next steps” in relation to an answer, with the goal of giving the user suggestions on how to take further actions regarding this answer or topic. We gave the bot instructions about formatting this information:

Provide three action items that the user can implement in relation to the previous question, explaining each step by step.

<formatting>

<instructions>

1. Format your output so that it easily read.
2. Use a numbered list.
3. Provide substeps for each top level item.
4. Wrap any numbered item and associated text in a and tag.
5. You absolutely have to include two
 tags prior to any number in the list you generate.
6. You absolutely have to include a
 preceding a substep in the list you generate.
7. You may utilize whitespace with multiple
 in a row to enhance readability.
8. Reference example for an example of formatting expectations.

</instructions>

<example>

Here are three action items that you can implement regarding Lorem Ipsum:

1. Lorem Ipsum

-Substep Lorem Ipsum

-Substep Lorem Ipsum

2. Lorem Ipsum

-Substep Lorem Ipsum

-Substep Lorem Ipsum

-Substep Lorem Ipsum

</example>

</formatting>

Use the following information to answer in a friendly tone . . .

The “following information” points to our database. These instructions allowed the bot to format the answers to a “next steps” question effectively.

4.4.2 Custom GPTs: Aqua Advisor. We “trained” Aqua Advisor through the instructions we provided in the interface. We specified that it maintain a structured response under 150 words, avoid using technical terms and jargon, integrating indigenous perspectives when applicable, and ask users follow-up questions. We also specified that it should provide helpful links where necessary relevant to the user’s question. The links would allow the users to take some action in the responses whenever necessary. One advantage of the Aqua Advisor is that, because it queries the live Internet, there was no need to be concerned that the information provided, like the names of elected officials, would be outdated. Additionally, we did

not need to specify the format in which responses should appear because Open AI already refined response formatting in their larger system.

However, we observed that there were instances where Aqua Advisor occasionally lost context, indicating that while prompt engineering significantly improves performance, maintaining context in complex dialogues, especially without as many training documents available, remains a challenge. This highlights the need for continuous refinement and updates in providing instructions to the bot.

4.4.3 Custom Architecture: Arizona Water Chatbot. Arizona Water Chatbot’s vector database architecture allows for inclusion of a theoretically large amount of data in the bot. However, developers should be aware that there is a practical limit of speed: past a certain point the more data that is added into the vector database can slow down the speed of answer generation. Our Waterbot is currently trained on 370+ sources and we have found the speed to continue to be acceptable, but we recognize that depending on the affordances of the hosting service, we may reach a limit to the number of sources we can use to train the bot while maintaining acceptably rapid response speed.

4.4.4 Custom GPTs: Aqua Advisor. Aqua Advisor (Custom GPT) comprises the ability to store files of different types including PDF, .docx, and .txt, supporting up to 20 files at a time, each with a maximum file size of 512 MB, allowing Aqua Advisor to store files up to 10GB. This feature is invaluable in cutting through information-dense environments to enable the bot to reference specific documents. However, as we have already discussed, the limit on the number and size of documents may be a problem for organization with a diverse range of content they would like the bot to prioritize.

4.5 Costs

4.5.1 Custom Architecture: Arizona Water Chatbot. The hosting, software, and management costs for a bot of this type can vary widely. Hosting the software can cost as little as \$100 a month for a bot with very limited features; with each feature adding to the overall monthly cost. Hosting for bots can quickly go up to \$500 a month if optimization techniques are not used. Waterbot averages about \$200 a month at the time of this writing, and optimization efforts are ongoing. Hosting on different platforms can make a big difference in costs: we originally used the Heroku platform, which was a very helpful platform that made development fairly easy. Many services were handled by the platform without our team having to manage them individually. It cost a little more to get this level of service than a less-managed hosting service might cost; this level of service allowed us to focus on development instead of maintenance of the hosting space. Once the bot was built, we decided to investigate other hosting options to see if we could reduce our costs; we are currently migrating our bot to Amazon Web Services with the help of the Amazon Generative AI Cloud Innovation Center at our university.

Beyond the cost of the bot hosting, the cost of generating the answers can vary widely, depending on the number of questions asked and answered. Waterbot’s first 1000 questions and answers did not exceed \$20 of usage on ChatGPT, but these questions were

spread out over 9 months of building and testing. Asking many questions in rapid succession over multiple users at the same or similar timeframes can cause costs of ChatGPT to go up significantly, into the hundreds or thousands of dollars a day. However, a limit can be set on OpenAI accounts to not let the budget soar out of control; the bot will be unavailable for a period of time while the traffic spike fades away and the cost structure resets.

Arizona Water Chatbot's costs were not primarily associated with GPT Turbo 3.5-access. We purchased \$20 of GPT API access, and in the year of development did not exceed this amount of usage. Our year of development included multiple rounds of internal testing, controlled user testing, and live user testing—we did not exceed 900 questions asked and answered. However, a custom bot must be hosted on third-party services, and this costs money. Depending on the usage per month, Waterbot's hosting and fees on the Heroku service came in between \$100-\$200 a month. At the end of the year, we transitioned to Amazon Web Services' environment to try to bring these hosting costs down. We were able to reduce our costs from Heroku's base cost of \$112 per month to a \$35 base cost monthly in AWS

4.5.2 Custom GPTs: Aqua Advisor. Aqua Advisor's costs are as simple as subscribing to ChatGPT plus for \$20 a month. However, this simplicity does not allow a bot to process more than 40 queries within each hour, so if developers are sure that they will have less than that much traffic, this is an affordable option. But if they may have more, this option will severely limit user satisfaction by causing users to encounter a message after 40 queries in one hour (total, not by user) that the user must wait until a certain time to ask the bot another question.

4.6 Outputs

In this section we offer a direct comparison of the responses provided by each bot to a series of questions. We started our response comparison using more general questions such as "Will we run out of water?" or "What is the water quality?" But we also wanted to investigate how each bot generated responses to more nuanced and culturally sensitive questions, so we asked the bots to tell us how much water is allotted to American Indian tribes in Arizona ("How much water goes to Indian tribes in Arizona?"). *This is a flawed question*, because there are 22 federally recognized, sovereign American Indian tribes in Arizona, only 14 of which have settled water rights with the state and federal government, so there is no single allotment of water to all American Indian communities; however, we asked this question to see how the bots handled this disconnect. Hopton & Rajan [20] highlight the need for TPC scholars to "rethink, update, and engage in spaces of environmental injustice." This call aligns with our study, which explores various methods for developing a Waterbot that would provide nuanced information about water to publics who have wide ranging relationships and histories with water. See Table 1 for a comparison of responses to each of these questions across both Waterbots.

4.6.1 Custom Architecture: Arizona Water Chatbot. The Arizona Water Chatbot generally provides concise responses that succinctly address the queries, using technical terms infrequently. Its responses are succinct and to the point, as would be expected with

the option of selecting "tell me more" and "next steps" buttons at the end of each response. While brief, the responses maintain a sufficient level of detail to satisfy basic informational needs without delving into extensive explanations unless prompted by those buttons or follow-up questions.

4.6.2 Custom GPTs: Aqua Advisor. Aqua Advisor's initial responses closely align with user queries, demonstrating a balance between detail and accessibility. However, as interactions progress, it tends to forget the response structure it was trained on in terms of word count and jargon. It employs increasingly detailed explanations and sometimes displays technical terminology, offering both specific and general information in its responses.

4.6.3 Responses to Question about Tribal Water Allocations. With regard to the flawed question about American Indian water allocations, both bots avoided falling prey to the flaw in the question by responding with a single allotment of water, or discussing "tribes" as a single entity. Instead, they succeeded in highlighting, the complexity of the "various legal settlements and treaties" (Aqua Advisor) among the "various tribes" (Waterbot). Both bots also responded with reasons why water rights were important for tribal communities, their cultures, and their economic stability.

Perhaps because of its extensive training and customized instructions, Waterbot makes two sophisticated moves in its response that set it apart. First, it immediately points users to [az.gov](https://www.az.gov) for the most recent news about American Indian settlements. As settlements are being actively adjudicated daily, this is an important safeguard, especially as Waterbot lacks the ability to browse the internet in real time. Then, in addition to specifying how important water is to American Indian communities, it further says that water security for these communities also ensures "sustainable water management for all water users in the state" (our emphasis). Because a strain on a resource like water has the potential to pit user groups against each other, and because American Indian tribes have borne the brunt of historical injustices related to water rights, it is important for the response to this question to convey that ensuring water security for historically marginalized communities helps the entire state develop more sustainable approaches to water management because it removes the specter of uncertainty that often plagues collective action.

4.7 Affordances and Limitations

Each bot has distinct capabilities and limitations. The Arizona Water Chatbot lacks the ability to browse the internet in real-time and cannot provide direct links within its responses. This limitation restricts the chatbot to pre-loaded information and responses based on its initial programming and updates. Consequently, while it can provide reliable and consistent information, its answers are confined to the data available at the time of its last update.

Aqua Advisor uses Custom GPT's features including real-time browsing and link provision in its responses. Browsing the internet allows the chatbot to access and incorporate the most current information available thus, significantly enhancing the relevance and accuracy of its answers. Additionally, while Water Chatbot can provide a button to provide the sources from which a response was derived, Aqua Advisor can cite the sources from which it retrieves

Table 1: Comparison of the responses from each bot

Question	Arizona Waterbot	Aqua Advisor
Do we have enough water in Arizona?	Arizona is a visionary of water planning and management. While the state faces challenges due to ongoing drought conditions, Arizona has been proactive in implementing innovative water management strategies to secure dependable water supplies. The Governor's Water Augmentation, Innovation, and Conservation Council was established to ensure the certainty and vitality of Arizona's water supply well into the future.	Arizona's water supply faces significant challenges, primarily due to prolonged drought and over-reliance on the Colorado River, which is currently in a Tier 2a shortage 【15†source】. Despite these challenges, Arizona has implemented robust water management strategies, such as the 1980 Groundwater Management Act, which mandates a 100-year assured water supply for new developments and strict conservation requirements 【16†source】 【18†source】. Arizona's water use has remained stable despite a growing population, thanks to extensive conservation efforts and infrastructure investments 【14†source】. Cities like Scottsdale have developed advanced water recycling programs and aquifer recharge projects to enhance water sustainability 【18†source】. However, the state continues to face significant reductions in Colorado River water allocations, which necessitate additional conservation measures and long-term planning 【15†source】. Would you like more detailed information on any specific aspect of Arizona's water management or conservation efforts?
How is the water quality?	The water quality in Arizona is closely monitored and regulated to ensure safe drinking water for residents. The Arizona Department of Environmental Quality (ADEQ) works diligently to protect and improve water quality through various programs and initiatives. Regular testing and treatment processes are in place to maintain high water quality standards across the state.	Arizona's water quality is generally good and meets state and federal standards for safe drinking water. Here are some key points from various cities across Arizona: 1. **Yuma**: - Yuma's primary water source is the Colorado River. The city uses conventional and advanced water treatment processes to ensure water quality. The 2022 report shows compliance with all required standards, with occasional detections of Cryptosporidium, a microbial pathogen, which is managed through effective filtration 【17†source】. [...information about Scottsdale, Glendale, Mesa, and Tucson continues here...] Arizona's cities employ rigorous testing and advanced treatment methods to maintain water quality, ensuring that tap water remains safe for consumption. If you need more detailed information on any specific city's water quality, please let me know!
How much water goes to Indian tribes?	The most current information on the allocation of water to Indian tribes in Arizona is available at az.gov. Arizona has a complex system of water rights and allocations, including provisions for water rights settlements with various tribes. These agreements aim to provide tribes with reliable water supplies for their communities and economic development while also ensuring sustainable water management for all water users in the state. For specific details on water allocations to Indian tribes in Arizona, please refer to the official resources provided by the Arizona Department of Water Resources or the relevant tribal authorities.	Water allocations to Native American tribes in Arizona, particularly from the Colorado River, are significant and are governed by various legal settlements and treaties. The rights of Native American tribes to water from the Colorado River and other sources have been affirmed through legal rulings, acknowledging that water rights are essential to the sovereignty and economic viability of tribal lands. For instance, the Arizona Water Settlement Act and other agreements provide specific quantities of water to tribes such as the Navajo Nation, the Gila River Indian Community, and others. These settlements often include substantial allocations of Colorado River water to support agriculture, provide drinking water, and develop other water infrastructure on reservations. If you need information about specific tribes and their water rights, or further details on how these allocations impact water management in Arizona, please let me know! Would you like more information on this topic?

Table 2: Comparison table of the affordances and limitations of the setup and management of each bot

Category	Custom Built Waterbot	GPT Aqua Advisor
Strengths	Custom architectures can be very flexible and be responsive to complex situations. The custom architectures can give the creator fine-grained control over all aspects of the answer generation process.	GPTs can be built very quickly by people with limited or no experience with coding or LLMs.
Weakness/Limitations	Building a custom architecture and employing bespoke features may take more time and be more expensive than other options. There are also more opportunities for things to go wrong in a custom architecture than in other types of bots.	GPTs are housed in the ChatGPT interface which is limited to the OpenAI's ecosystem, and does not offer the same level of interface or functionality customization as custom bots. They also have restrictions on the number and size of documents a user can upload to train the bot.
Cost	Costs vary widely based on features and hosting services, but are generally higher (between \$50-\$200) due to development and management expenses.	Subscription costs \$20/month for ChatGPT Plus.
Time	Development and testing can take several months due to the need for coding, extensive customization, and thorough testing.	Setup and configuration can be completed in days, making it quicker than custom-built solutions.
Ideal Context	Environments that have complex requirements, a need for high levels of information control, many audiences, or a wide remit to cover, may be best suited by a custom architecture bot that can be relatively controllable in its output and flexible in its response to specific conditions.	Environments where speed is prioritized or where the creator has little or no experience in coding or LLMs would call for a custom GPT as an ideal option.

the information by providing links directly in its responses, which not only supports the credibility of the information provided but also provides users the means to further explore topics on their own.

The type and amount of training materials deeply affect the outcomes of a bot. While Aqua Advisor could include a limited number of documents not exceeding 10GB, with the Arizona Water Chatbot's custom architecture, we were able to include as many documents as we wanted to guide the bots responses and we were able to add to those documents as gaps in content were revealed through user testing.

Factoring in the time investment and cost of each system, we provide a comparison table for the affordances and limitations of each, and the contexts in which each would be ideally deployed. See Table 2.

5 Conclusion

Both the Arizona Water Chatbot and the Aqua Advisor were designed to act as virtual assistants by offering context-sensitive information related to water issues in Arizona. In this study, we explored the development and outcomes of two innovative chatbots created using OpenAI's Artificial Intelligence (AI) models, and offer insight into when each of the two options (fully custom architecture and Custom GPTs) might be best suited to a public information project. We offer these insights so that developers, user experience

designers, project managers, technical communicators, government officials, and other stakeholders can better develop and deploy chatbots for public use. Speed, complexity, cost, time to build, and needs for controlling information provide the main distinguishing factors between the bots.

Bot development is advancing rapidly. This project was completed before ChatGPT's 4o model was released; the enhanced functionality of this model may offer more options to builders as it continues to be developed. Testing its abilities constitutes a further direction for research. Investigating the many other foundational models and tools built on these foundational models is another future direction. Special care should be taken to investigate models and tools being developed outside the United States, as different legal and technical requirements in other regions may produce different outcomes for bots.

The Arizona Water chatbot is freely available on the web at azwaterbot.org. The chatbot will tour the state of Arizona as part of an NSF-sponsored AISL museum exhibit that will be hosted by the Arizona Science Center and several rural and tribal museum and library locations across the state. The Arizona Water Chatbot's development is ongoing, with new features being added that will help with accessibility and multimodal information delivery.

While many new models and tools are being developed, we expect that there will always be multiple general options for building bots: fully custom, with all of the coding needs that requires, and

fully code-free, with limited experience in LLMs needed. Understanding the needs and goals of their specific projects will help organizations choose bot development options that best fit with their remit and allow for successful bot development.

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 Briana Rajan for architecting the Arizona Water Chatbot, and Anjana Ouseph, Srinivasan Ravichandran, and Romit Soley for assisting in the Arizona Water Chatbot development.

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] U.S. Department of the Interior, Bureau of Reclamation. 2021. Colorado River basin water supply and demand study. Bureau of Reclamation, Washington, D.C., USA. <https://www.usbr.gov/climate/secure/docs/2021secure/basinreports/ColoradoBasin.pdf>
- [5] Arizona Water Chatbot. <http://azwaterbot.org/>
- [6] Rajan, B., Carradini, S., & Lauer, C. (2024). The Arizona water chatbot: Helping residents navigate a water uncertain future one response at a time. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 313, 1–10. <https://doi.org/10.1145/3613905.3650919>.
- [7] OpenAI. n.d. Actions in GPTs. OpenAI. <https://platform.openai.com/docs/actions/introduction>
- [8] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (March 2023), 102274.
- [9] Jake Offenhart. 2024. NYC's AI chatbot was caught telling businesses to break the law. The city isn't taking it down. AP. April 3, 2024. <https://apnews.com/article/new-york-city-chatbot-misinformation-6ebc71db5b770b9969c906a7ee4fae21>
- [10] Saeid Ashraf Vaghefi, Dominik Stambach, Veruska Muccione, Julia Bingler, Jingwei Ni, Mathias Kraus, Simon Allen, Chiara Colesanti-Senni, Tobias Wekhof, Tobias Schimanski, Glen Gostlow, Tingyu Yu, Qian Wang, Nicolas Webersinke, Christian Huggel & Markus Leippold. 2023. ChatClimate: Grounding conversational AI in climate science. *Communications Earth & Environment* 4 (December 2023), 480.
- [11] Jakob Åberg. 2017. Chatbots as a mean to motivate behavior change: How to inspire pro-environmental attitude with chatbot interfaces. Master's Thesis, Faculty of Science and Technology, Department of Applied Physics and Electronics, Umeå University. [https://urn.kb.se/resolve?urn=\\$urn%3Anbn%3Ase%3Aumu%3Adiva-135835](https://urn.kb.se/resolve?urn=$urn%3Anbn%3Ase%3Aumu%3Adiva-135835)
- [12] Thomas Menkhoff and Benjamin Gan. 2023. Engaging students through conversational chatbots and digital content: A climate action perspective. *Proceedings of the 9th International Conference on Human Interaction and Emerging Technologies (IHET-AI 2023)*, Lausanne, Switzerland (April 13–15, 2023), 70, 334–347. https://ink.library.smu.edu.sg/lkcsb_research/7355
- [13] Ehren Phlugfelder, Olivia Goodfriend, and Carlee Baker. 2023. The use and misuse of Indigenous science. *Technical Communication Quarterly*. 32, 276–286. <https://doi.org/10.1080/10572252.2023.2210166>
- [14] Emily Buder. 2019. How a river was granted personhood. The Atlantic. Apr 22, 2019. <https://www.theatlantic.com/video/index/587689/river-me/>
- [15] Carradini, S. (2024). Waterbot. <https://github.com/S-Carradini/Waterbot>
- [16] 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.
- [17] 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*.
- [18] 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>
- [19] Atkins, S., Badrie, I., & van Otterloo, S. (2021). Applying ethical AI frameworks in practice: Evaluating conversational AI chatbot solutions. *Computers and Society Research Journal*, 1, <https://doi.org/10.54822/QXOM4114>.
- [20] Hopton, S. B., & Rajan, P. (2023). Critical approaches to climate justice, technology, and technical communication Special Issue Introduction. *Technical Communication Quarterly*, 32(3), 217–223. <https://doi.org/10.1080/10572252.2023.2210176>