



# @Ilegra: a chatbot for Vallader

Oliver Bendel<sup>1</sup> · Dalil Jabou<sup>1</sup>

Received: 2 November 2023 / Accepted: 2 February 2024 / Published online: 19 February 2024  
© The Author(s) 2024

**Abstract** Extinct and endangered languages have been preserved primarily through audio conservation and the collection and digitization of scripts and have been promoted through targeted language acquisition efforts. Another possibility would be to build conversational agents like chatbots or voice assistants that can master these languages. This would provide an artificial, active conversational partner which has knowledge of the vocabulary and grammar and allows one to learn with it in a different way. The chatbot, @Ilegra, with which one can communicate in the Rhaeto-Romanic idiom Vallader was developed in 2023 based on GPT-4. It can process and output text and has voice output. It was additionally equipped with a manually created knowledge base. After laying the conceptual groundwork, this paper presents the preparation and implementation of the project. In addition, it summarizes the tests that native speakers conducted with the chatbot. A critical discussion elaborates advantages and disadvantages. @Ilegra could be a new tool for teaching and learning Vallader in a memorable and entertaining way through dialog. It not only masters the idiom, but also has extensive knowledge about the Lower Engadine, that is, the area where Vallader is spoken. In conclusion, it is argued that conversational agents are an innovative approach to promoting and preserving languages.

**Keywords** Chatbot · Language model · GPT · Vallader

## Abbreviations

AI Artificial intelligence  
API Application Programming Interface

GPT Generative Pre-Trained Transformer  
HTTP Hypertext Transfer Protocol  
LLM Large language model  
RTR Radiotelevision Svizra Rumantscha  
SRF Schweizer Radio und Fernsehen/Swiss Radio and Television  
TTS Text-to-speech  
UI User interface  
UNESCO United Nations Educational, Scientific and Cultural Organization

## 1 Introduction

Language is one of the most fundamental of human characteristics. Many animals also have one form of it, but humans can use it to conquer the world and build a civilization. Individual languages are cultural tools and cultural assets [1]. It is usually our concern to promote and preserve them, whether for functional or aesthetic reasons. Many languages are doing well—they are in demand and widely used. They exist in spoken and, more recently in human history, in written form. English has had a unique success story, dominating vast areas of the world—particularly, the USA, Canada, Australia, and the UK—and of science. Chinese and Indian variants are of great importance simply because of their population size and the economic power of these countries [2].

Other languages are doing less well. They are threatened or endangered, as in the case of the Sardinian or Welsh languages or Sanskrit [3]—or even completely extinct as in the case of the Egyptian. They are neither in demand nor used on a large scale. Nevertheless, they can be very important for individuals, groups, and societies and not only from functional and aesthetic perspectives. Language is also home, is containment and demarcation to persons and groups: it

✉ Oliver Bendel  
oliver.bendel@fhnw.ch

<sup>1</sup> School of Business FHNW, 5210 Windisch, Switzerland

creates identity. The loss of a language can have serious consequences. With it, a cultural tool is lost as well as a cultural asset. At worst, a whole culture can disappear. According to the linguistic anthology “Ethnologue”, which regularly compiles an index of the most and least spoken languages, over 42% of the world’s languages are endangered [2]. According to the UNESCO Atlas of Endangered Languages (“UNESCO” stands for “United Nations Educational, Scientific and Cultural Organization”), a distinction is made between “safe”, “vulnerable”, “definitely endangered”, “severely endangered”, “critically endangered”, and “extinct” [4].

The main author of this paper has been developing conversational agents in the context of artificial intelligence (AI), machine ethics, and social robotics at his university since 2012 [5–9]. This has involved giving them moral rules or letting them show empathy and emotion to help people. In 2022, another focus was added. The lead author and his rotating teams have been trying to use chatbots and voice assistants to rescue and promote endangered and dead languages. The idea is that schools and universities, as well as the general public, will be provided with such dialog systems. The users—i.e., the pupils and students—should be able to text or speak with them and thus learn a language or stay in practice. One could see the project in the context of “AI for Good” or “AI for the Good” [10].

The first project, which began in mid-2022 and ended in early 2023, was devoted to Latin, a so-called dead language. This is still taught at schools and universities and is required in some places for admission to courses such as philosophy. @ve was created as an artificial interlocutor, with text input and output. The second project (described in this paper) began in early 2023 and ended in August 2023 [11]. It was dedicated to an endangered language. A chatbot for the Rhaeto-Romanic idiom Vallader was developed. It masters text input and text output and has a voice output. It was named @llegra after the Rhaeto-Romanic word for “Hello!”.

New language models permitted new possibilities for the development of chatbots. The lead author had been working on GPT-2 (“GPT” stands for “Generative Pre-Trained Transformer”) since 2019 and had included a colleague’s contribution on the use of the large language model (LLM) in social robots in his edited volume [8]. GPT-3 had already been published and was known for its increased effectiveness. With the help of @ve, experience with it could be gained. While the project was still in progress, the even more powerful GPT-4 was released, which proved to be a stroke of luck for the implementation of @llegra [11].

This paper starts with the basics of Rhaeto-Romanic and Vallader and explains basic terms like “chatbots”, “GPT-3”, and “GPT-4”. In each case, it places these in the project’s context. Then it presents the preparation and implementation of the @llegra project. It also shows the testing of the chatbot by several native speakers. A critical discussion of

the results follows. At the end there is a summary with an outlook.

## 2 Conceptual basics

As part of the conceptualization, the authors worked out the basics of the background and meaning of Vallader. They both live in Switzerland and were exposed to Rhaeto-Romanic, but were not experts in the fourth national language, as it is often called. They also worked on the basics of chatbots and GPT-4, also in connection with Rhaeto-Romanic. The main author had experience with LLMs since 2019, starting with GPT-2.

### 2.1 Background and importance of Vallader

Graubünden Romansh—often simply called Romansh or Rhaeto-Romanic—belongs to the Rhaeto-Romanic languages together with Dolomite Ladin and Friulian [12]. It is an official language in the canton of Graubünden (Grisons) in Switzerland, along with German and Italian. At the federal level, it is the fourth national language of Switzerland, along with German, French, and Italian. Nowadays, 36’217 adults (0.5% of the Swiss population) consider themselves Rhaeto-Romanic speakers and report it as their main language [13]. There are five idioms, i.e., varieties with a standardized written language: Surselvisch (Sursilvan), Sutselvisch (Sutsilvan), Surmeirisch (Surmiran), Oberengadinisch (Puter, Putèr), and Unterengadinisch (Vallader).

The Vallader idiom is spoken by an increasingly smaller community in the Lower Engadin between Martina and Zernez and in Val Müstair [14]. In recent decades, the number of speakers has drastically decreased and there is a risk that the language will die out in the near future [15]. As of 2023, it can be estimated that there are about 7000 Vallader speakers left. There are many overlaps with other idioms, but there are also many words of their own. For Vallader, there are lexicons, school textbooks, and children’s books, as well as literary works. Radio and television stations are also available. Radiotelevision Svizra Rumantscha (RTR), the only Romansh media house (<https://www.rtr.ch>), says it uses all five Romansh idioms and Rumantsch Grischun.

The federal government and the canton of Graubünden are making various efforts to promote the Romansh idioms. This includes familiarizing schoolchildren with the basics. A standard language has also been introduced, Rumantsch Grischun [15]. However, this is neither liked nor actively spoken by the population—it is a language of the authorities, but also a way to subtitle films and broadcasts. Most recently, digitization measures have been planned and implemented, such as educational videos for children. A chatbot or voice assistant would be an important addition to this.

The focus should be on the five idioms, and it is important to have at least a voice output in addition to text input and text output so that students and other learners can hear and mimic the language.

## 2.2 Chatbots as conversational agents

Chatbots can be traced back to Alan Turing or Joseph Weizenbaum, depending on the narrative [16–18]. They are dialog systems with natural language capabilities of a textual or auditory nature [19, 20]. They are used, often in combination with avatars, on websites or in instant messaging systems, where they explain and advertise the products and services of their operators or take care of the concerns of interested parties and customers [21]—or simply serve for amusement and reflection. Some chatbots act as virtual learning companions [22]. Sometimes the term chatbots is so broadly defined that it also includes voice assistants with voice input and output. Chatbots can be rule-based and access a knowledge base with statements and answers or they can be located in the field of artificial intelligence [23] and use methods such as machine learning. These can also be termed as conversational agents or conversational AI [24, 25].

Chatbots of today are often available in English or Chinese. This covers numerous countries and language communities. Japanese, German, French, and Spanish chatbots are also available. European users particularly appreciate being able to communicate in the language of their country. Spanish and French, moreover, are world languages. Chatbots have also been developed for Arabic [26]. Since chatbots are primarily used for business and entertainment, endangered or extinct languages have not been of interest so far. Development of a classical rule-based chatbot would also have been too time-consuming and cost-intensive. Machine learning and especially language models offer more powerful and, at the same time, simpler possibilities here [27].

## 2.3 The language models GPT-3 and GPT-4

GPT-3 was introduced in June 2020 and is the third in a series of language models capable of generating intelligible and meaningful text [28]. It is based on the Transformer architecture and has been trained by generative pre-training with a very large amount of data [29]. GPT-3 differs from other models mainly in its size and performance capabilities. With 175 billion parameters—also called neurons—it was the largest language model to date until the release of GPT-4 and comparable models. The transformer architecture in combination with the amount of pre-trained data enables GPT-3 to solve complex tasks in the field of Natural Language Processing (NLP) that were previously impossible or difficult for smaller models.

GPT-4 is the latest version of the GPT models developed by OpenAI and was released on March 14, 2023. It represents an improvement over previous models in the series in terms of various performance metrics [30]. According to OpenAI [31], GPT-4 is a multimodal model that accepts image and text input and produces text output [31]. While the model performs less well than a human in many real-world scenarios, it does demonstrate human-level performance on various professional and academic benchmarks. Nevertheless, this should not tempt us to confuse it with human reasoning.

The biggest improvement of GPT-4 over GPT-3 is the ability to process more complex tasks with improved accuracy. This enables a wide range of applications suitable for everyday use, from dialog systems to speech generators and robotic applications. The biggest difference between GPT-3 and GPT-4 can be seen in the number of parameters. While GPT-3 was trained with 175 billion parameters, as mentioned earlier, making it the largest language model created to date, GPT-4 was trained with approximately 100 trillion parameters (according to other sources it was less) [30].

The Chat Completion API (“API” stands for “Application Programming Interfaces”) was developed specifically for the newer GPT-4 applications that are optimized for conversational interfaces [11]. Unlike the Completion API, which expects only text input, the Chat Completion API uses a special transcript format inspired by chats. The models were developed with a “conversation-in, message-out” approach and can be used in both multi-level conversational contexts and non-chat scenarios. A simple API call for a chatbot like @llegra might look like Fig. 1.

Central in this context is the “messages” parameter [11]. This must contain an array of objects, each of which has a specific role (“system”, “user”, or “assistant”) and content. The “system” messages are used to control the behavior of the assistant. Here, the personality of the chatbot can be defined or specific instructions for its behavior during the conversation can be specified. The “user” messages contain requests or comments to which the assistant should respond. The “assistant” messages store previous responses of the assistant and can be used as examples of desired behavior [32].

The large amount of data also includes Rhaeto-Romanic texts. Apparently, OpenAI has scanned entire books—in principle, encyclopedias, textbooks, children’s books, and literary works are available. Social media could also be relied upon in this case, or chat conversations. Throughout Switzerland, it is common to text in the respective dialect or idiom using instant messengers, with some people using a translation service such as Textshuttle to do so [33]. However, as conversations have revealed, in Graubünden it can be a mixture of idioms, Rumantsch Grischun, German and Italian words, and emojis (e.g., of a farmer or a Capricorn).

**Fig. 1** Example for a chat completion API



```

import openai

openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "Du bist ein hilfreicher Assistent."},
        {"role": "user", "content": "Was ist die Hauptstadt von Graubünden?"},
        {"role": "assistant", "content": "Die Hauptstadt von Graubünden ist Chur"}
    ]
)

```

The language model GPT-4 does not allow the user to train it with further data, such as teaching materials for Vallader. In the future, however, such opportunities will exist.

### 3 Preparation of the @llegra project

The @llegra project was initiated in early 2023 at the School of Business FHNW by the main author. He also acted as the client, with the corresponding rights (use and further development of the prototype) and duties (provision of a budget, in this case 500 Swiss francs, i.e., about 500 US dollars). He was able to collaborate with the co-author for the implementation [11]. Until then, the developer had no knowledge of Rhaeto-Romanic and, above all, the motivation to take a closer look at GPT and possible applications.

The chatbot was named @llegra at the suggestion of the main author, after the Rhaeto-Romanic word for “Hello” or “Happy”. The name could also be that of a futuristic girl or woman. As will be seen, an avatar was chosen for @llegra that suited its cultural context. The main author was available to provide expertise to the co-author throughout the project. Technical details were also discussed with an additional supervisor, Bradley Richards. The project lasted half a year, from February to August 2023.

At the beginning of the project, it was still unclear which idiom should be chosen. Vallader was on the shortlist because it was an idiom that had enough speakers, but not too many, and was definitely considered threatened. Another candidate was the widely spoken idiom Sursilvan. A lucky coincidence was that a text-to-speech (TTS) engine for Vallader was found, making this the obvious choice.

#### 3.1 First steps with GPT-3

Based on the experience with @ve, it was clear that @llegra should also be based on GPT [34]. Accordingly, the project was set up and work began on feeding GPT-3 with Romansh

texts. The Lia Rumantscha, the umbrella organization of all Romansh language associations, provided a collection of children’s books. The developer used four of them in the project. After fine-tuning the GPT-3 model, he began testing. This was done using PyCharm (<https://www.jetbrains.com/de-de/pycharm/>), an integrated development environment.

Initially, it was verified that the language model understands the zero-shot approach (where only a natural language description of the task is provided and the model predicts the answer) and provides appropriate answers [11]. However, this approach was not very successful. The language model did not respond like a chatbot and generated confusing or incoherent responses to questions in Vallader. As a result, the few-shot approach was tested. The model was given a few examples to improve its comprehension ability. With this approach, it was able to understand most of the language and occasionally even provided answers in Romansh. However, these were also incoherent and out of context.

During preliminary work on the project, OpenAI released GPT-4 [31]. After extensively studying the results and limitations of the GPT-3 model, the team decided to make a switch to this newer version of the language model [11]. This provided an opportunity to overcome the existing challenges and limitations while potentially achieving better performance and more accurate responses.

#### 3.2 Replacement by GPT-4

Before the evaluation of the GPT-4 model could begin, it was necessary to apply for additional developer rights from OpenAI. Due to a high rush in spring 2023, it was necessary to wait approximately three weeks before access was granted. Once access was obtained, testing of the GPT-4 model could begin. Similar to the handling of the GPT-3 model, the GPT-4 model was evaluated directly in PyCharm.

The “system” messages are of central importance in this API call, as they define the role of the chatbot and guide its behavior. Accordingly, the developer experimented with

commands that define goals and tasks. To give the assistant a more precise idea of how it should react to questions, he added a sample conversation (with the messages “user” and “assistant”).

Surprisingly, GPT-4 was already able to respond in Rhaeto-Romanic during the first interaction without extensive text input. Various test runs showed that the chatbot almost always responded in Rhaeto-Romanic: the answers provided to posed questions were given in this language and made sense in terms of content. With the help of the described messages, the accuracy improved further in the test phase. In addition, it turned out that the chatbot has profound knowledge about the Lower Engadine.

To ensure that the answers were actually in Vallader or at least in Rhaeto-Romanic, the answers were translated in each case using the Textshuttle tool mentioned earlier. After several more tests and adjustments to the system message, the team decided to use the GPT-4 model for the final deployment in the chatbot.

## 4 Implementation of @llegra

The chatbot consists of a server file and a client file. The server file contains the connection to the OpenAI API and the implementation of the TTS engine, while the client file provides the user interface (UI). The user interface also includes the avatar. The choice fell on an image of an ibex (in English and Rhaeto-Romanic also known as “Capricorn”) since one appears on the coat of arms of Grisons and is the symbolic animal of the canton. The image was generated with the image generator Gencraft (<https://gencraft.com>), with the prompt “A majestic ibex to us as an avatar”. Since @llegra is reminiscent of a female name, she was repeatedly referred to as Capricorn girl by the lead author. In the next sections, the most important functions of the respective files are explained in more detail. Then the integration into the website is described.

### 4.1 Server-side implementation

In order to ensure that the chatbot is able to read out its answers, research was already carried out during the preparation of the project to find an appropriate TTS engine. In the process, the team members discovered the solution from SlowSoft (<https://www.slowsoft.ch>). This company was founded in November 2014 by two experts in the field of speech technology and specializes in the development of speech synthesis. Between 2016 and 2017, the experts worked on the creation of speech data for Swiss German and Rhaeto-Romanic (Vallader). SlowSoft offers a range of products that can be used to implement speech synthesis in chatbot [35]. It is a speciality that the company also

offers solutions for dialects and idioms. The website states: “Today we offer speech synthesis products in Swiss German (Grisons dialect and Zurich dialect), Rhaeto-Romanic (Vallader idiom), Swiss Standard German, US English.”

Contacting the company allowed the team to investigate the best way to integrate their engine with the chatbot. Two options were available: either connect via an API or deploy the engine directly. In order to develop a seamless and independent chatbot, direct deployment was preferred. This decision increases the complexity of the development but guarantees the high quality of the output. SlowSoft created the required files for local execution and delivered them to the team.

As part of the implementation of the interaction with the OpenAI API, a knowledge base was first loaded. An essential element here, as already explained, is the definition of the “system” message. With this, the personality and role of the chatbot can be defined. The statement here is: “Du bist ein hilfreicher Assistent, der darauf trainiert ist, Fragen auf Vallader zu beantworten. Du kannst Inputs auf Vallader, Deutsch oder Englisch nehmen, aber deine Antworten sind immer auf Vallader.” (“You are a helpful assistant trained to answer questions on Vallader. You can take inputs in Vallader, German, or English, but your answers will always be in Vallader”, translation by the authors) (Fig. 2). The instruction is, as said, the result of several tests.

The knowledge base also contains essential information and sample dialogs that help the chatbot better master the Vallader idiom and respond to queries more efficiently and effectively. The messages used here also come under “user” and “assistant”. The examples below illustrate how the user presents different sentences and dialogs in Vallader to the chatbot and how the chatbot responds to them (Fig. 3). Likewise, passages from the children’s books have been integrated, each preceded by a short explanation to help the bot understand the text more deeply. Such complementary measures help to improve the accuracy of the chatbot’s responses.

### 4.2 Client-side implementation

This section focuses on the client file in which the user interface of the chatbot is implemented [11]. The file is written in JavaScript using the React library (react.dev) and additionally employs the @chatscope/chat-ui-kit-react package to design the chat interface. To send HTTP requests (“HTTP” stands for “Hypertext Transfer Protocol”), the axios package (<https://axios-http.com>) is used.

A detailed description of the import of resources and packets, the initialization of state variables, message processing and dispatch to the server, client-side processing and display of the server response, the audio playback function of the TTS engine, and the main UI code is omitted here. Instead, the focus is on the “return” statement, which defines

```

OpenAI API

app.post('/processMessage', async (req, res) => {
  const chatMessages = req.body.chatMessages;

  let apiMessages = chatMessages.map((messageObject) => {
    let role = "";
    if (messageObject.sender === "ChatGPT") {
      role = "assistant"
    } else {
      role = "user"
    }
    return { role: role, content: messageObject.message }
  });

  const systemMessage = {
    role: "system",
    content: "Du bist ein hilfreicher Assistent, der darauf trainiert ist, Fragen auf
    Vallader zu beantworten. Du kannst Inputs auf Vallader, Deutsch oder Englisch
    nehmen, aber deine Antworten sind immer auf Vallader."
  }

  const apiRequestBody = {
    "model": "gpt-4",
    "messages": [systemMessage,
      ...predefinedMessages,
      ...apiMessages]
  }
}

```

**Fig. 2** Definition of the “system” message

**Fig. 3** Extension of the knowl-  
edge base with messages

```

wissensDatenbank.json

[
  { "role": "user", "content": "Allegra" },
  { "role": "assistant", "content": "Allegra, co poss jau gidar" },
  { "role": "user", "content": "Be amo tschinch nots e lura n'haja anniversari" },
  { "role": "assistant", "content": "Faina üna festina?" },
  { "role": "user", "content": "Hier ein kurzer Text, damit du besser Vallader kannst. Der
  Text ist ein Dialog und es geht um Zahnschmerzen: «Auaa», plondscha Lupo, «meis daints fan
  suul mal.» Lupo schmetta da rusigliar vi da seis öss e til lascha crodar our d'bocca. «Ha!
  N'ha bain pensà cha teis daints gnian marschs cun magliar glatsch! Uossa hast la merda,
  uossa varast il prüm dolurs, lura crodarà oura ün daint, lura amo ün e lura amo ün - fin
  cha tü nun hast plü ingüns. E lura poust be amo magliar schoppa e mösas fin cha tü
  mourast!», disch Brutus cun tun dad important. «Ojee, ma quai nu lessa!», disch Lupo e
  guarda tuot sculozzà." },
  { "role": "assistant", "content": "Grazia fitg per questa decleraziun" },
]

```

the presentation of the user interface. Here, various components from the chat UI kit library are used to create the chat window, the message list, and the input field (Jabou 2023):

- The main container (“MainContainer”) includes the chat container (“ChatContainer”). This is the basic structure of the chat window and its functions.
- Inside the ChatContainer, there is the message list component (“MessageList”). This contains all messages that are exchanged between user and bot. The array of messages is traversed, and a message component is rendered for each one.
- The direction of the message (“incoming” or “outgoing”) is determined based on the sender of the message (chatbot or user) and displayed on the left or right of the window accordingly.
- If the sender of the message is “Allegra” (an arbitrarily chosen placeholder), an “avatar” component is also added, which displays the Capricorn girl. Clicking on the avatar will replay the last reply.
- The “MessageList” component also has a “TypingIndicator” associated with it, which becomes visible when the bot is currently generating a response. This lets the user know that they will receive the response soon.
- Below the “MessageList” is the input field (“Message-Input”). When the user enters and sends a message, the “handleSend” function is called.
- Lastly, as mentioned earlier, there is a checkbox that allows the user to turn the text-to-speech function on or off. One can also use a speech output but does not have to. It is a prerequisite for replaying the message.

This code ultimately creates a functional chat application that enables dynamic interaction with the chatbot and provides a responsive user interface.

### 4.3 Website creation

The creation of the website was quite complex, especially on the server side [11]. After the project was successfully deposited on the Git repository, different server hosting platforms were compared. The choice fell on Render (<https://render.com>). There, the Git repository was connected and the API key for the OpenAI API was added as an environmental variable. Starting up the server initially went without error, but problems arose when the connection to the client was tested. It turned out that the synthesizer file of the TTS engine, which was developed for macOS, was not compatible with the Linux-based render platform. Therefore, the text-to-speech engine files had to be further customized. Since errors were still occurring, an open-source software platform called Docker (<https://www.docker.com>), which

allows applications to be created, deployed and run in containers, was used on Render’s recommendation.

Unlike server hosting, hosting the client was a relatively straightforward affair. The client was implemented using Netlify (<https://www.netlify.com>), a platform designed specifically for hosting and deploying front-end applications. First, the client Git repository was linked to Netlify, which allowed for a seamless transition and continuous integration. Once a change was made in the Git repository, Netlify was able to automatically capture the changes and deploy the latest version of the client.

An important aspect that had to be considered before uploading the client was to adjust the server address in the client code. The client must know where the server is hosted in order to send requests to it correctly. In addition, it was decided early on that the domain for the website should be `allegra-bot.ch`. The domain could be created and linked directly through Netlify.

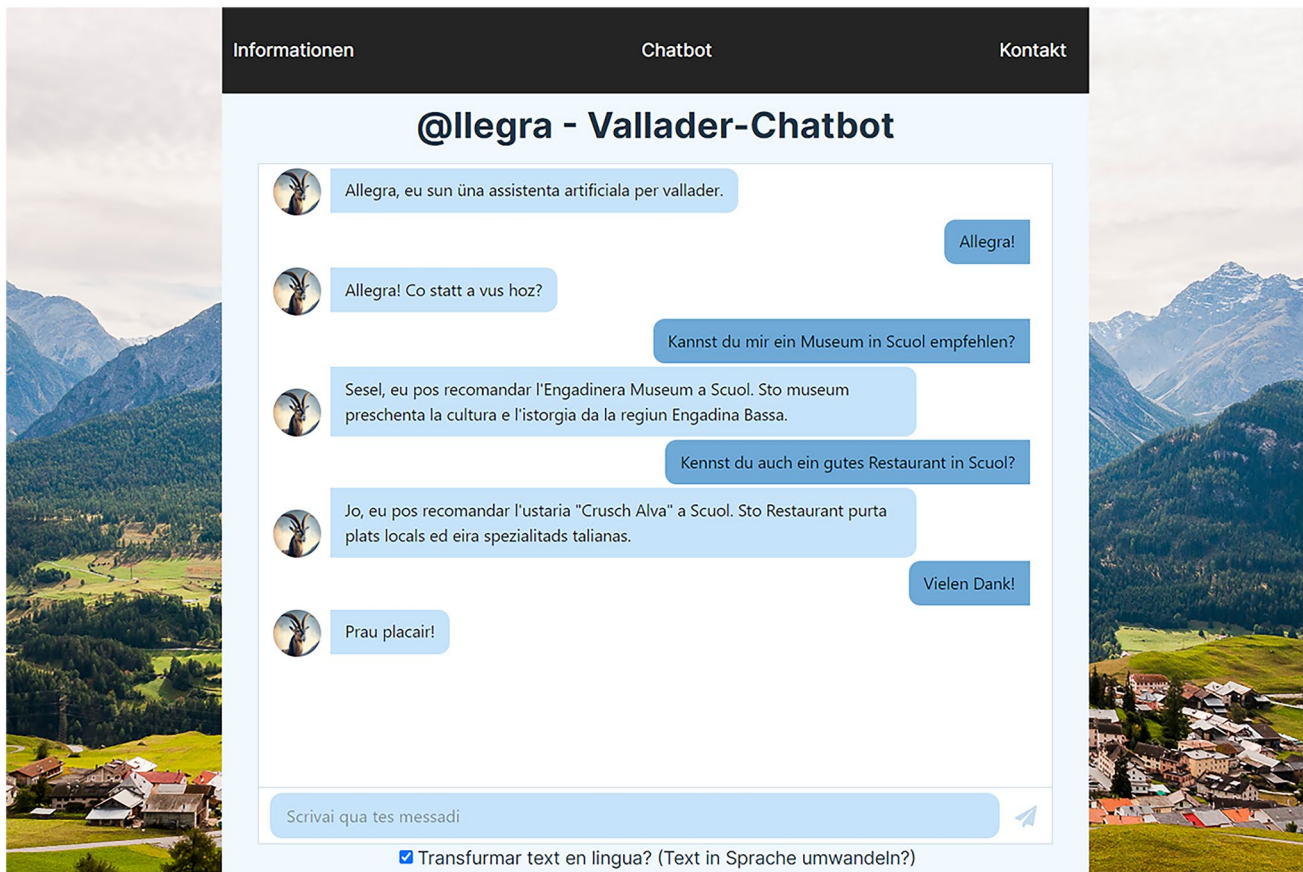
Once both the server and the client were successfully uploaded and hosted, user testing could begin, which will be discussed in the next chapter. Hosting the client was thus a critical step in the process, as it allowed the final product to be seen and tested in a live environment. In doing so, one was completely independent of the lab and could acquire participants from the general public.

The website itself comes across as uncluttered (see Fig. 4). There is a main navigation at the top: “Informationen” (“Information”), “Chatbot”, and “Kontakt” (“Contact”). The information page is displayed by default. It contains information about the project and the restrictions that the chatbot has. For example, it reads: “It should be noted that there may be a short delay in generating the first response when using the bot for the first time. However, once connected, responses should be generated relatively smoothly.” (translation by the authors). Clicking on “Chatbot” brings up @llegra. “Contact” takes you to photos and contact information for the main author and co-author. The background is a photograph of the Lower Engadine.

The chat window constantly shows a circle with an avatar, the already mentioned Capricorn girl. Inputs and outputs are displayed linearly, with the most recent at the bottom. @llegra’s statements and answers are left-aligned and highlighted with light blue shading, while the user’s statements and questions are right-aligned and highlighted with medium blue shading. At the very bottom, there is the already described option to activate or deactivate the speech output. It is enabled by default. The chat window fills the whole middle part of the web page.

### 4.4 Native speaker tests of @llegra

Five user tests were conducted as part of this project to assess the functionality and usability of @llegra and identify



**Fig. 4** Website of @Ilegra with chatbot interface

potential areas for improvement [11]. Two of them were not native speakers—they were the main supervisor and the second supervisor themselves. They entered text and analyzed the answers using the translation tool Textshuttle [33]. Three other testers were native speakers. As it happens, all of them were women. In two of these three tests (tests 1 and 3), conducted by the co-author on June 25 and July 25, 2023, conversation histories were stored in an Excel spreadsheet for analysis and users were asked six questions in German about the chatbot (translation by the authors):

1. How user-friendly did you find the chatbot's interface?
2. Did the chatbot understand your queries correctly?
3. Were the chatbot's answers understandable and helpful?
4. How would you rate the quality of the chatbot's voice output? Was it clear and easy to understand?
5. Did you experience any problems or challenges during the test?
6. Do you have any further comments or feedback?

One user of this test was a Rhaeto-Romanic speaker, but not a Vallader speaker. The other was a Vallader speaker. The answers to these questions served as valuable feedback

to further optimize the user experience. Another very short test with a Vallader speaker was conducted in the Lower Engadine by the lead author (test 2). This was timed to take place between the other two, on July 9, 2023.

In general, the graduates of the Excel tests were very satisfied with the user interface. No problems arose in starting the application and in the technical flow of the conversations. However, one user had to switch to a smartphone because her notebook only displayed parts of the website. The response times were good, quite better than expected (remember the warning on the website). The chatbot understood the questions correctly. The answers were mostly understandable and helpful. The Rhaeto-Romanic speaker wished that the answer could be replayed. The Vallader speaker pointed out that the answers are not always only in Vallader, but repeatedly contain expressions in Sursilvan or Rumantsch Grischun. In addition, she recognized a grammatical error in the initial greeting. The idea of the chatbot was found to be good.

For the second test, the main author traveled to the Lower Engadine with @Ilegra on his smartphone. In Tarasp Fontana, a mountainside village near Scuol, he found a suitable test subject, a woman of about 78. This Vallader speaker

listened to the chatbot's answer to a question about Tarasp Fontana and began translating what was said into German without being asked. Subsequently, @llegra was asked if she knew Scuol—the main town in the neighborhood. The chatbot then gave a detailed answer in Vallader. The elderly lady confirmed that the pronunciation was good and the content understandable. She was impressed with the chatbot's performance and asked if it was artificial intelligence.

This live test provided a valuable opportunity to get direct feedback from a native Vallader speaker. She vigorously confirmed that @llegra's voice output was indeed Vallader, which the lead author considered a very positive assessment, while being aware that it was only a snapshot. In fact, the assessment is also put into perspective by the testimony of the Vallader spokesperson in the third test, who tested the chatbot in much greater detail. In addition to the test with the Vallader speaker, the lead author took the opportunity to ask @llegra in Scuol more questions, especially regarding the local Museum d'Engiadina Bassa and local restaurants. This provided further valuable data, experience, and insights for the continuous improvement and adaptation of the chatbot.

It can be argued that five tests are not enough to generalize the results. However, this may not have been the intention. On the one hand, the main goal was to show that the chatbot works in principle, that it understands and masters the idiom. On the other hand, with chatbots based on generative AI, it is hardly possible to generalize the results, as they are based on probabilities and are to some extent random. Of course, the authors have tried to limit the randomness with examples from children's books, but even that has its limits. Last but not least, it is very difficult to find Rhaeto-Romanic speakers, which is due to the fact that there are not many native speakers left.

## 5 Discussion of the results

@llegra arose from both effort and coincidences. First, the project failed in the first attempt starting in August 2022, but led to an interesting, new project: the chatbot @ve. Then the project threatened to fail again in early 2023—but the release of GPT-4 gave it a new boost. Added to this was the discovery of the text-to-speech engine and the support of SlowSoft. In the end, a powerful but not perfect chatbot was created. It impresses everyone who has ever heard or read Romansh. It impresses even native speakers—but they are also the ones who recognize its weaknesses.

The tests of the three Rhaeto-Romanic speakers led to small but important adjustments. The first tester suggested that @llegra should repeat the answer if desired. The developer implemented this immediately. If you click on the avatar in the improved version, the answer is spoken again, provided that the TTS engine is activated. This

can be repeated as often as desired. The new feature is explained on the introductory page. It was helpful during the second test in the Lower Engadine. The feedback of the third tester was also taken into account. Among other things, the entry greeting, the introduction of the avatar, was corrected immediately on July 25, 2023.

A weakness of @llegra, as the tests of the third user have shown, is that the answers are not always in Vallader but contain terms in Sursilvan or Rumantsch Grischun. If Vallader is to be learned correctly, this is definitely a problem. To solve it, there is a technical and a didactic approach. The technical one consists in training GPT-4 with texts in Vallader. It can be assumed that OpenAI will grant the option as with GPT-3. The didactic one is to make it clear to the student that @llegra repeatedly uses expressions of other idioms or standard language. This could be justified by the fact that, in day-to-day reality, speakers of dialects also acquire and use words of other dialects and languages, for example, if they have lived in other areas or have had frequent contact with their speakers. However, the Lia Rumantscha, for example, welcomes purity of the idiom, as conversations have shown.

The second test in the Lower Engadine produced the idea to use the chatbot not only for language maintenance, but also as an original tourist guide. For non-native speakers, it would certainly be necessary that the chatbot could then also output English, Italian, and German or other languages. You would kill two birds with one stone, so to speak: you would have a chatbot with a lot of knowledge about the region, about the history, about the museums, and about the restaurants. In addition, you would be able to hear the language of the region and perhaps remember at least one or two words. This would allow one to better connect with the population, which of course has always grown up at least bilingual. Moreover, one could show one's appreciation towards the language and culture of the Lower Engadine.

The Swiss media already reported on the project during the last project phase at the end of July. First in print media (21 July and 26 July 2023), then the local radio station, Schweizer Radio und Fernsehen/Swiss Radio and Television (SRF), in Grisons [36, 37]. Some of the contributions also mentioned their own tests and the address of the website with the chatbot. This resulted in a considerable increase in the number of users. A limit of \$120 per month was set. OpenAI did not respond to several requests to increase the limit to \$300. So, a page had to be put up stating that the chatbot was temporarily unavailable. Then, as of July 28, 2023, @llegra was ready to talk again. For cost reasons, the chatbot was temporarily shut down again on August 11. Trial access could be granted upon request. The chatbot was freely available again from September

2023. Basically, one challenge was that this is only a low-budget project.

## 6 Summary and outlook

This contribution to the project on the chatbot @llegra first established basics on Rhaeto-Romanic, and specifically on Vallader, and explained basic terms such as “chatbots”, “GPT-3”, and “GPT-4” within the project’s context. Then, it presented the preparation and implementation of the project, up to its integration into the website. In addition, the tests of the chatbot by three native speakers were discussed. This was followed by a critical discussion of the results.

@llegra was the first chatbot based on GPT to be created for Vallader. This could support an endangered language in that users can communicate with her and can hear the language too. Teaching and learning materials on Vallader usually do not have this interactive component. Thus, the language is practiced and learned in a very limited way. The project’s findings could also be applied to chatbots for extinct languages, for example.

With @llegra, a powerful chatbot was created that goes far beyond the capabilities of @ve. This also speaks to the suitability of GPT-4 for such projects. However, several difficulties also came to light. For example, the chatbot was not able to speak pure Vallader. Probably the written base was too small for that after all, or it could not always correctly distinguish between the idioms and the standard language during operation. It also became apparent that it can be difficult to persuade an American company to be flexible and raise the user limit. This also limited the opportunity to show the chatbot to as many users as possible and get feedback from them.

There are several interesting possibilities for future projects or further developments of @llegra. One of them is the integration of more Rhaeto-Romanic idioms into the chatbot, which could be achieved through a kind of selection box where users can choose the idiom in which they want to receive answers. This would provide greater variety and flexibility for users and could help increase awareness and appreciation of the different Rhaeto-Romanic idioms. However, this would require the development of additional text-to-speech engines for the respective idioms, which is certainly possible thanks to the progress made in this field.

**Acknowledgements** The authors would like to thank the three external testers. They want to remain anonymous. As native speakers, they were a great help in evaluating the capabilities of the chatbot. In addition, the authors would like to thank Simonne Bosiers and Nadine Ganz from the University of Applied Sciences of the Grisons for helpful information and for providing access to a corpus of texts in Vallader.

**Author contributions** OB: developed the theoretical framework for the project and formulated the basic requirements. DL: programmed the chatbot. Both authors tested the chatbot with several people. They also read and approved the final manuscript.

**Funding** Open access funding provided by FHNW University of Applied Sciences and Arts Northwestern Switzerland. The submitted work was not supported by any organization.

**Data availability** The knowledge base containing the Vallader datasets is archived at <https://github.com/oliverbendel/allegra>. The full answers from the testers can also be found there [38].

### Declarations

**Conflict of interest** We declare that we do not have any interests that are directly or indirectly related to the work submitted for publication.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bendel O (2023) Sprache. In: Gabler Wirtschaftslexikon. <https://wirtschaftslexikon.gabler.de/definition/sprache-124739>. Accessed: 1 December 2023
- Eberhard DM, Simons GF, Fennig CD (ed) (2023) Ethnologue: Languages of the World, Twenty-sixth edition. <http://www.ethnologue.com>. Accessed: 1 December 2023
- Koul N (2021) Manvi SS (2021) A proposed model for neural machine translation of Sanskrit into English. *Int j inf tecnol* 13:375–381. <https://doi.org/10.1007/s41870-019-00340-8>
- Moseley C, Alexandre N (2010) Atlas of the World’s languages in danger. VW ‘s-Hertogenbosch, Hertogenbosch
- Anderson M, Anderson SL (2011) Machine ethics. Cambridge University Press, Cambridge
- Bendel O (2018) From GOODBOT to BESTBOT. In: The 2018 AAAI Spring Symposium Series, Palo Alto. pp 2–9
- Bendel O (ed) (2019) Handbuch Maschinenethik. Springer VS, Wiesbaden
- Bendel O (ed) (2020) Maschinenliebe: Liebespuppen und Sexroboter aus technischer, psychologischer und philosophischer Sicht. Springer Gabler, Wiesbaden
- Bendel O (ed) (2021) Soziale Roboter: Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen. Springer Gabler, Wiesbaden
- Vieweg SH (ed) (2021) AI for the Good: Artificial Intelligence and Ethics. Springer, Cham

11. Jabou D (2023) @Ilegra—Chatbot für Vallader: Ein Conversational Agent für eine gefährdete Sprache. Bachelor Thesis, School of Business FHNW.
12. Liver R (2012) Rätoromanisch, Historisches Lexikon der Schweiz, HLS. <https://hls-dhs-dss.ch/de/articles/024594/2012-06-19>. Accessed: 1 December 2023
13. Bundesamt für Statistik (2023) Hauptsprachen seit 1910: Ständige Wohnbevölkerung ab 15 Jahren. Hauptsprachen seit 1910—1910–2021. <https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/sprachen-religionen/sprachen.html>. Accessed: 1 December 2023
14. Grison (w.d.) Grison Vallader. <https://www.grison.ch/sprache/vallader>. Accessed: 1 December 2023
15. Deflorin T (2023) Rätoromanische Sprache, Radiotelevision Svizra Rumantscha (RTR), 11 April 2023. <https://www.rtr.ch/emissions/decodar-nossa-cultura/raetoromanisch/fakten-gesichte/fakten-und-zahlen-raetoromanische-sprache>. Accessed: 1 December 2023
16. Weizenbaum J (1966) ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM* 9(1):36–45
17. Shah H, Warwick K, Vallverdú J, Wu D (2016) Can machines talk? Comparison of Eliza with modern dialogue systems. *Comput Hum Behav* 58:278–295
18. Shum H, He X, Li D (2018) From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Front Inf Technol Electron Eng* 19(1):10–26
19. Kohne K, Kleinmanns P, Rolf C, Beck M (2020) Chatbots: Aufbau und Anwendungsmöglichkeiten von autonomen Sprachassistenten. Springer Vieweg, Wiesbaden
20. Bendel O (2017) Chatbot. In: Gabler Wirtschaftslexikon. <http://wirtschaftslexikon.gabler.de/Definition/chatbot.html>. Accessed: 1 December 2023
21. Pandey S, Sharma S, Wazir S (2022) Mental healthcare chatbot based on natural language processing and deep learning approaches: Ted the therapist. *Int j inf tecnol* 14:3757–3766. <https://doi.org/10.1007/s41870-022-00999-6>
22. Sakulwichitsintu S (2023) ParichartBOT: a chatbot for automatic answering for postgraduate students of an open university. *Int j inf tecnol* 15:1387–1397. <https://doi.org/10.1007/s41870-023-01176-z>
23. Chowdhary KR (2020) Fundamentals of Artificial Intelligence. Springer, New Delhi
24. McTear M (2020) Conversational AI: dialogue systems, conversational agents, and chatbots. Springer International Publishing, Cham
25. Bendel O (2023) Conversational Agent. In: Gabler Wirtschaftslexikon. <https://wirtschaftslexikon.gabler.de/definition/conversational-agent-125248>. Accessed: 1 December 2023.
26. Youness F, Madkour MA, Elshenawy A (2023) Dialog generation for Arabic chatbot. *Int j inf tecnol*. <https://doi.org/10.1007/s41870-023-01519-w>
27. Maree M, Al-Qasem R, Tantour B (2023) Transforming legal text interactions: leveraging natural language processing and large language models for legal support in Palestinian cooperatives. *Int j inf Tecnol*. <https://doi.org/10.1007/s41870-023-01584-1>
28. Dale R (2020) GPT-3: what's it good for? *Nat Lang Eng* 27(1):113–118
29. Gillioz A, Casas J, Mugellini E, Khaled OA (2020) Overview of the Transformer-based Models for NLP Tasks. In: 15th Conference on Computer Science and Information Systems (FedCSIS), Sofia, Bulgaria. p 179–183
30. Weißgraeber R (2023) GPT-4: All You Need to Know + Differences To GPT-3 & ChatGPT. *AX Semantics*, 13 March 2023. <https://en.ax-semantics.com/blog/gpt-4-and-whats-different-from-gpt-3/>. Accessed: 1 December 2023
31. OpenAI (2023) GPT-4, 14 March 2023. <https://openai.com/research/gpt-4>. Accessed: 1 December 2023
32. OpenAI Platform (w.d.) Welcome to the OpenAI developer platform. <https://platform.openai.com>. Accessed: 1 December 2023
33. Schüssler M (2023) Wie gut übersetzt Textshuttle Schweizerdeutsch und Rätoromanisch? *Tagesanzeiger*, 25 May 2023. <https://www.tagesanzeiger.ch/wie-gut-uebersetzt-textshuttle-schweizerdeutsch-und-raetoromanisch-605292991507>. Accessed: 1 December 2023
34. N'diaye K (2023) @ve: Ein Chatbot in der Sprache der alten Römer. Bachelor Thesis, School of Business FHNW.
35. SlowSoft (w.d.) Die Schweizer Stimme für Ihr Produkt. <https://www.slowsoft.ch/deu/index.html>. Accessed: 1 December 2023
36. Laukenmann J (2023) Eine KI zur Rettung des Rätoromanischen. *Tagesanzeiger*, 26 July 2023. <https://www.tagesanzeiger.ch/eine-ki-zur-rettung-des-raetoromanischen-221418099510>. Accessed: 1 December 2023
37. Schwendener T (2023) Wie GPT-4 das Rätoromanische retten soll. *Inside IT*, 21 July 2023. <https://www.inside-it.ch/wie-gpt-4-das-raetoromanische-retten-soll-20230721>. Accessed: 1 December 2023
38. Github (w.d.) Data and documents for the @Ilegra project. <https://github.com/oliverbendel/allegra>. Accessed: 1 December 2023