computers

Article

Impact of Chatbots on User Experience and Data Quality on Citizen Science Platforms

Akasha-Leonie Kessel [1,*], Soror Sahri [2], Sven Groppe [1,*], Jinghua Groppe [1], Hanieh Khorashadizadeh [1], Marc Pignal [3], Eva Perez Pimparé [4] and Régine Vignes-Lebbe [3,5]

[1] Institut für Informationssysteme, Universität zu Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; jinghua.groppe@uni-luebeck.de (J.G.); hanieh.khorashadizadeh@uni-luebeck.de (H.K.)
[2] Laboratoire d'Informatique Paris Descartes, Université Paris Cité, 45 Rue des Saints-Pères, 75006 Paris, France; soror.sahri@u-paris.fr
[3] Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, 57 Rue Cuvier, 75005 Paris, France; marc.pignal@mnhn.fr (M.P.); regine.vignes-lebbe@mnhn.fr (R.V.-L.)
[4] Infrastructure Récolnat, Direction Générale Déléguée Aux Collections, Muséum National d'Histoire Naturelle, 57 Rue Cuvier, 75005 Paris, France; eva.perez@mnhn.fr
[5] Laboratoire Informatique et Systématique, Sorbonne Université, 1 Rue Victor Cousin, 75005 Paris, France
* Correspondence: akashaleonie.kessel@student.uni-luebeck.de (A.-L.K.); sven.groppe@uni-luebeck.de (S.G.)

Abstract: Citizen science (CS) projects, which engage the general public in scientific research, often face challenges in ensuring high-quality data collection and maintaining user engagement. Recent advancements in Large Language Models (LLMs) present a promising solution by providing automated, real-time assistance to users, reducing the need for extensive human intervention, and offering instant support. The CS project Les Herbonautes, dedicated to mass digitization of the French National Herbarium, serves as a case study for this paper, which details the development and evaluation of a network of open source LLM agents to assist users during data collection. The research involved the review of related work, stakeholder meetings with the Muséum National d'Histoire Naturelle, and user and context analyses to formalize system requirements. With these, a prototype with a user interface in the form of a chatbot was designed and implemented using LangGraph, and afterward evaluated through expert evaluation to assess its effect on usability and user experience (UX). The findings indicate that such a chatbot can enhance UX and improve data quality by guiding users and providing immediate feedback. However, limitations due to the non-deterministic nature of LLMs exist, suggesting that workflows must be carefully designed to mitigate potential errors and ensure reliable performance.

Keywords: Large Language Model (LLM); LLM application; chatbot; user interface; citizen science; data quality

check for updates

Academic Editor: Ming Liu

Received: 9 November 2024
Revised: 3 January 2025
Accepted: 7 January 2025
Published: 10 January 2025

Citation: Kessel, A.-L.; Sahri, S.; Groppe, S.; Groppe, J.; Khorashadizadeh, H.; Pignal, M.; Perez Pimparé, E.; Vignes-Lebbe, R. Impact of Chatbots on User Experience and Data Quality on Citizen Science Platforms. *Computers* **2025**, *14*, 21. https://doi.org/10.3390/computers14010021

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Citizen science (CS) is an approach that allows the general public to participate in scientific research, allowing non-experts to contribute to data collection, analysis, and interpretation [1]. By engaging the general public, CS projects can leverage a large pool of volunteers to gather valuable data in vast geographic areas and over long periods of time. This participation can significantly reduce the resources needed for scientific projects, as it complements the efforts of professional researchers with the resources of motivated citizens. Hence, CS is seen as key for current and future research initiatives in national strategies [1].

This paper focuses on a specific use case: the "Les Herbonautes" project. This CS initiative involves participants in the digitization and analysis of botanical specimens [2].

The project relies heavily on the contributions of non-expert volunteers to collect data on plant samples, which necessitates robust data validation strategies to maintain data quality.

A critical aspect of CS, especially for data collection projects, is ensuring the high quality of the data collected [3]. Since participants are not always experts in the field, the accuracy and reliability of the data they provide can vary widely. To address this challenge, many CS projects employ various data validation mechanisms, like cross-validation, where at least two users have to enter the same input, expert or peer verification, where a user enters input which an expert or a different user then controls, or they employ automated systems that control user input according to predefined rules [3,4]. Depending on the strategy used, a large amount of monetary and time-related resources is necessary (i.e., expert verification) [5]. Unlike existing work, we introduce a system to significantly reduce the resources needed for traditional data validation strategies by combining different strategies that check user input in real time and that can be used not only for data input, but also for onboarding, or project-related questions, and input recommendations. To achieve this, we leverage the capabilities of Large Language Models (LLMs), which have become pervasive in applications like copilots, assistants, translators, and reviewers. LLMs offer advanced natural language understanding and generation capabilities, making them highly suitable for tasks requiring dynamic, user-adaptive solutions. Contributions like [6] envision LLMs as key components in data pipelines. Building on these concepts, we present one of the first practical realizations of an LLM-based system focused specifically on improving data quality in the data collection tasks, specifically in the CS project "Les Herbonautes". Our proposed system includes a user interface (UI) in the form of a chatbot that can interact with users in natural language. Using the LangGraph framework, a network of LLM agents was created that successively solves specific subtasks. Unlike existing approaches, this graph-like architecture ensures that the system can handle queries with complex task structures while maintaining the quality of the output through integrated feedback loops.

By integrating this network of LLM agents into the data collection workflow, our system improves both data quality and user experience (UX). The user-centric design, incorporating a user-friendly interface, real-time feedback, and onboarding support, encourages more accurate and efficient data input, which directly impacts data quality. This dual focus addresses general challenges in data collection while offering a tailored solution for "Les Herbonautes". This work makes novel contributions, combining LLM-driven interactions with UX improvements to advance data quality assurance in CS. While similar works target a specific step in CS projects, our solution assists projects in multiple steps of data collection and user engagement. We used the LangGraph framework to create a network of LLM agents that is able to solve even complex data validation and user interaction tasks in the field of Herbarium Label Identification. This allows for a modular architecture and separately improvable workflows, which makes enhancing and abstracting the system very simple. The product uses an LLM-powered chatbot as an interface for data input, question answering, and recommendations which have never been combined in this context before.

In the following sections, we initially conduct a comprehensive review of the existing literature to identify related work. We then analyze the specific context of use of the "Les Herbonautes" project, including exchanges with project administrators, which helped gather additional information on areas for potential improvements to the current system. Combining this information with insights from the literature, we develop a set of formal requirements to guide the design and implementation process. Based on these requirements, we conceptualize and implement a solution as a functional system. The implemented system is evaluated by assessing its UX and usability.

2. Related Work

This section reviews existing research and practices in key areas related to our work, highlighting the gaps our contribution aims to address. We organize it into three main areas: data validation strategies in CS projects, the integration of LLM agents to solve multiple tasks related to CS, and the use of LLMs specifically in biodiversity platforms. Our literature review emphasizes the shift between traditional validation methods, such as peer and expert verification, and the emerging role of LLM-powered chatbots in improving data validation processes. In the following, we explore relevant methods and their impact on data quality, emphasizing their potential and limitations in the context of CS projects.

2.1. Data Validation in CS Projects

Data quality is a critical concern in CS, necessitating robust mechanisms for both data verification and data validation. While data verification is primarily focused on ensuring the accuracy of the identification of the recorded items, data validation involves standardized, often automated checks to ensure the completeness, accuracy of transmission, and validity of the record's content [3]. Our work focuses mainly on data validation, aiming to develop and refine methods that enhance the reliability of data collected through CS projects.

Within the scope of CS projects, it is evident that data control mechanisms are a widely recognized and implemented practice. Of the 103 CS projects considered in the study [4] and dedicated to identifying alien species with the help of citizen scientists, 91 have established validation procedures. Data uploaded to CS portals can undergo various types of validation procedures. We distinguish between four primary approaches to validating CS data: peer verification, expert verification, automatic quality assessment, and model-based quality assessment [3]. The most common methods for validating CS data are peer verification and expert verification, both of which are frequently supported by automated filtering techniques [4].

2.1.1. Peer and Expert Verification

In the context of CS projects, peer verification generally involves citizen scientists who partake in the data collection process validating the records submitted by other citizen scientists. This approach leverages the collective knowledge and expertise within the community, so the overall quality of the data heavily depends on the active engagement and expertise of the citizen scientist community involved in the project [4]. Expert verification, much like peer verification, is a widely adopted approach in many CS projects. This method involves specific contributors, classified as experts, reviewing and verifying the data submitted by citizen scientists [4]. Expert verification adds a layer of professional oversight to the data validation process, which can enhance the credibility of the findings derived from CS projects.

Peer and expert verifications require human reviewers to assess the quality and accuracy of the data provided by citizen scientists. Thus, this process can result in delayed feedback, thereby limiting the opportunity for immediate data correction and reducing the responsiveness of the validation process. Many CS portals employ a combination of expert verification and other validation methods to optimize data quality [4].

2.1.2. Automatic Quality Assessment

CS projects have increasingly started to incorporate technologies to enhance the efficiency and accuracy of their data collection processes. One such advancement involves the use of software-based systems designed to automatically assess the quality of the generated data. These systems ensure that the contributions made by volunteers, often varying widely in expertise, are both accurate and reliable [4].

Due to the current growth of machine learning approaches in many different fields, the integration of Artificial Intelligence has become a significant trend in the evolution of CS projects [7]. AI technologies, particularly machine learning models and natural language processing (NLP) tools, offer robust solutions for handling large datasets and complex information, which are common in many CS initiatives.

Automated quality assessments often occur in real time applying predefined rules or algorithms to detect anomalies or outliers. This automatic validation not only reduces the need for extensive human oversight, thereby saving resources, but also provides instant feedback options to participants [4].

2.1.3. Model-Based Quality Assessment

Model-based quality assessment represents an advanced approach to data validation in CS projects, building upon the capabilities of automatic quality assessment systems. Model-based quality assessment employs sophisticated models that are specifically trained to evaluate the quality of data according to the unique requirements of each project. Unlike basic automated systems, model-based quality assessment requires input and guidance from domain experts to effectively train the system for a specific use case [4].

2.2. Utilising LLMs as Chatbots

In the development of chatbots, a significant challenge has been their historical reliance on simple state machine architectures. These early designs, while functional, severely limited the chatbot's ability to engage in dynamic conversations. Users were often unable to deviate from a predetermined conversation flow, resulting in rigid interactions that failed to accommodate more complex or natural user inputs [8]. This is why, so far, the utilization of chatbot systems in CS projects has been limited to question answering and onboarding tasks. With the creation of chatbots that build on LLMs like ChatGPT 4, users are now able to interact with chatbots in natural language and dynamically, ensuring the inclusion of context information. Due to this improvement, chatbot systems experienced a rise in usage and might now be applied to a broader and more complex range of contexts.

Effective CA design is highly dependent on the domain for which it is built [9]. In data collection contexts, the chatbot must be tailored to handle specific types of queries and data formats relevant to the project's goals. A well-designed CA can navigate the complexities of the domain, guiding users through the data entry process with domain-specific prompts and checks that ensure high data quality [9].

Additionally, by employing NLP and understanding the capabilities of LLMs [10], chatbots can ask follow-up questions, provide clarifications, and encourage more detailed responses, thereby reducing the likelihood of superficial answers [11].

As users interact with a chatbot in a conversational style, they may perceive the interaction as a form of social engagement rather than a mechanical task. This shift in perception can make the task feel less burdensome and more like a meaningful exchange, thereby increasing user motivation and reducing the cognitive load associated with completing the task [11]. However, this enhanced engagement also entails significant ethical considerations. Transparency becomes a critical issue in these interactions.

The impersonation of human agents by AI systems may manipulate users' emotions or decisions in unethical ways. Biases inherent in training datasets can amplify stereotypes [12]. Additionally, AI systems are prone to hallucination, where they generate incorrect information with high confidence [13]. The problem is intensified when users are unaware of the AI's limitations or the probabilistic nature of its outputs. Thus, users should know whether they are interacting with an AI system or a real human at all times, and to what extent their data is being used, so they can adequately decide on their next

action. To raise trust and accountability, it is important to implement clear AI disclosure policies and ensure that users are aware of the nature of their interactions.

Providing real-time feedback and maintaining an ongoing dialogue helps to maintain user focus and attention, ensuring that the data collected is of high quality [11,14].

The use of LLMs as chatbots in data collection projects offers significant advantages by automating processes that previously required substantial human resources. CAs, powered by LLMs, provide a low-threshold solution that can lead to substantial cost reductions [14]. By automating repetitive and labor-intensive tasks, such as data entry preliminary data validation, or question answering, these systems reduce the need for extensive human intervention, allowing resources to be allocated more efficiently [5].

Once the data in CS projects are collected, they might be used to improve the LLM the chatbot is based on. For example, the study in [15] presents a case study about fine-tuning an LLM for nutrition advice based on citizen-generated data. Other approaches [16] for creating CAs avoid relying on fine-tuning by using techniques such as few-shot prompting [17], chain-of-thought [18], and external memory.

Quality Assessment of Chatbots

When trying to assess the quality of chatbots and conversational agents (CAs), a significant focus lies in identifying key quality attributes that can gauge their performance. These attributes are crucial for determining how well chatbots meet user expectations, particularly in terms of usability. These quality attributes align closely with the ISO 9241 concept of usability, which emphasizes "the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments". This alignment suggests that principles traditionally applied to software usability are equally relevant when evaluating chatbots, albeit with considerations unique to these interactive systems [19].

A challenge in assessing chatbot quality is the difficulty in establishing a universal scale. Given the diverse range of applications for chatbots and CAs—from customer service to entertainment—different use cases may prioritize different aspects of quality, such as responsiveness, accuracy, or empathy [19]. To address these complexities, it is useful to employ methodologies that enable the direct comparison of chatbot systems [19].

Studies in customer service settings have highlighted key factors that influence user satisfaction with chatbots, such as problem resolution and precision and concreteness of responses [14,20]. These studies suggest that, when chatbots provide clear, accurate, and actionable answers, user satisfaction increases significantly. In contrast, errors and lack of functionality can quickly lead to user frustration and distrust [20]. These findings are directly applicable to the design of chatbots for scientific and data collection purposes, where the clarity and precision of the information are critical to maintaining user engagement and trust [11].

2.3. LLMs in Biodiversity Platforms

LLMs are increasingly being integrated into biodiversity platforms. LLM-powered chatbots can vary significantly in complexity, ranging from simple conversational units that handle basic inquiries [21] to more advanced systems that integrate NLP, machine learning, and data mining technologies. Depending on the complexity, these systems are able to do anything from holding simple conversations with users to assisting them during specific workflows. More sophisticated systems are often well suited to a wide range of CS projects, where the ability to process and understand complex data is essential [4].

As CS projects grow and attract more participants, the demand for expert feedback and guidance increases. Traditional methods, relying heavily on human experts, are often insufficient to meet this demand. By deploying LLM-based chatbots, platforms can scale

their operations more effectively, providing immediate feedback of high quality to a large number of users without the need for an equivalent increase in expert resources [4].

However, the responses of chatbots need to be critically analyzed. The contribution in [22] includes such an analysis of AI-generated responses from the ChatGPT chatbot for a question interview about ecological restoration. It focuses on the distributive, recognition, procedural, and epistemic justice dimensions in ChatGPT's answers about ecological restoration and identifies issues with organizational, gender, geographical representation, and diverse knowledge systems.

Research on frameworks to develop task-oriented LLM-powered chatbots designed to perform specific tasks relevant for the biodiversity domain just started [23].

For example, the authors of [24] compiled a question-and-answer test set based on Darwin Core records available in the Integrated Digitized Biocollections (iDigBio) (available at https://www.idigbio.org/ (accessed on 13 December 2024)) and presented the result of testing ChatGPT's biodiversity knowledge with it. In [25], the same authors extend their contributions by training a model to predict the ability of GPT-4 to reproduce iDigBio species occurrence data, which improves species absence and presence predictions and allows extrapolation beyond the records available in iDigBio to visualize geographic distributions of individual species.

As an alternative approach, the authors of [26] applied Monte Carlo Dropout (MCD) and Expected Calibration Error (ECE) to evaluate the uncertainty of LLMs for establishing reliable generative question-answering models with the use case of biodiversity conservation. Finally, the authors of [27] report on the integration of a chatbot into the CS project iDigBio and its use cases to perform searches, including visualizing the geographic distribution of found records on an interactive map and packaging the results of the searches as a Darwin Core Archive to be emailed to the user, and to collect statistics.

The authors of [28] propose an application that leverages multimodal LLMs for the identification of fungus species and their edibility based on images uploaded by users. Furthermore, the application provides educational material on the risks associated with consuming the identified fungus, cooking recipes for edible fungi, and ecological information.

In [7], the authors' objective was to create a system able to transcribe the information from herbarium labels as accurately as possible using LLMs. They tested different LLMs to analyze the possibilities that LLMs in biodiversity research might have while focusing mainly on the transcription process. The system was implemented on a separate platform and does not employ a chatbot for user interaction.

In biodiversity platforms, human experts are often in short supply, and the pre-prepared materials available for providing feedback to participants are inherently limited in scope. In such contexts, the use of chatbots containing a Natural Language Generation (NLG) component can significantly enhance the richness and relevance of feedback provided to users [29].

This type of rich context-aware feedback not only helps improve the accuracy of future identifications but also enhances the educational value of the platform, fostering a deeper understanding of biodiversity among participants [4].

Overall, there are only a few existing works [27,28] that utilize LLM-based chatbots in CS projects in the domain of biodiversity, and one about advanced natural language text generation for user feedback (without using LLMs) [29]. Table 1 compares the existing projects to our solution, showing a unique set of supported features, design, and focus in our solution.

Table 1. Comparing existing work with our proposed approach.

	Focus			Application Functionality								Design		
Approach	**Citizen Science**	**Data Quality**	**Support of Data Validation**	**Personalizable Interaction**	**Enabling Database Queries**	**Onboarding**	**Question Answering**	**Considering Platform-Specific Info**	**Considering Web Search Results**	**OCR Integration**	**Image Recognition Integration**	**Agentic Workflow**	**Open Source LLMs**	**No Data Transfer to Third Parties**
NLG4Nature Conservation (Bumblebee) [29]	✓	✓	✗	✗	✗	✓	✗	✓	✗	✗	✓	✗	✗	✓
Fungi identification [28]	✗	✓	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗
iDigBio chatbot prototype [27]	✓	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗
Voucher Vision prototype [7]	✓	✓	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗
Our Approach	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓

In the following paragraphs, some of the table's features will be explained in more detail.

The first three columns describe the focus of the related projects. The focus is very similar among all projects, as we chose to compare our work with solutions that were integrated into similar contexts. Data validation strategies, which are used after the input of data from citizen scientists to improve data quality by controlling the validity of the input, are not part of any of the reviewed projects. The next group of columns describes the application functionalities present in the systems. As our approach saves completed missions and botanically relevant interests the users give to the system during interaction with it, it can personalize the recommendation of further missions by analyzing this information and ranking the existing missions. More information about missions and the general workflow of our system is given in the next sections. When users have plant-specific questions, about for example its already recorded collectors, our system can generate MySQL queries to retrieve information from the mission's database. Other projects simply do not employ project-related databases or do not allow for queries to them. For other questions that are more general, an LLM agent can decide to run a web search on it so that the system can answer the user's question. Related systems do not include web search components. We differentiate between OCR integration and Image Recognition Integration. For herbarium records, the integration of OCR means that the system would be able to analyze the text on the labels, which could help to acquire information about the discovery. Image Recognition in this context takes into account the entire image, focusing more on the details of a specimen in the hopes of gaining information on the species or particular traits of the specimen that are prevalent in the image. Since for our project the species of the specimen has already been recorded, we focus on analyzing the label information by using OCR algorithms.

The last group of columns gives information on the architecture of the different systems. Our solution is built on a network of LLM agents that split tasks into subtasks which are then processed by single agents, creating a workflow able to solve even complex tasks. Since the LLMs used for the agents are open source, the system can be implemented locally. If so, there is no need for data transfers to third parties. As soon as any models with costs are used, which is the case for almost all the other projects, data is transferred to other

countries and companies. The Bumblebee project [29] built its own NLG component and thus does not send data to third parties.

3. A Citizen Science Use Case: Les Herbonautes

The Les Herbonautes project is a digitization project by the Muséum National d'Histoire Naturelle (MNHN) from recent years, during which all 6,000,000 specimens of vascular plants and macroalgae, which were so far only stored analogously, have been digitized [2].

High-quality photos of those plant samples and additional information from the attached labels have been taken and stored online at http://coldb.mnhn.fr (accessed on 8 January 2025). However, this does not signify the end of digitization efforts, as digitizing the information on the attached labels remains a problem.

After uploading the pictures in 2012, the Les Herbonautes website (http://lesherbonautes.mnhn.fr, accessed on 13 December 2024) was created as a CS platform to analyze and digitize the information at hand. To guarantee high-quality data, so far, different data validation steps have been established in the project [2]. The first is the partition of information potentially available on the labels and photos into different skill groups. Since there are seven fields, seven skill levels have been implemented on the project's website. For every sample, any new user can fill in the first field of information, which is the country of origin. To level up and thus be able to input more complex information, users need to contribute to the current field a certain number of times for different sample images, in the hopes of improving their expertise while doing so. If a user has contributed to a particular field enough times to reach a higher level, they are asked to partake in a quiz providing guidelines on how to transcribe information from the next field. This is supposed to assist the user in building the competencies needed to correctly transcribe information classified as more complex. After data input from the user, the answers are validated further by using cross-validation. For each field of information, at least two contributors have to enter the same data for the field to be validated. For more complex fields, an administrator may decide to increase the number of identical inputs that are needed to validate the field. The last step implemented to guarantee high-quality data is through a chat forum. For every sample, contributors can post and answer questions about the data collection there. Since the forum is built like a group chat and supervised by an experienced botanist, users are encouraged to interact with others to receive reliable help. This not only benefits the quality of the collected data but also improves the feeling of solidarity between the users of the platform who are called herbonauts.

In 2016, the average number of contributions was 35,000 per month, with most of them being made by occasional visitors or by a small group of experts and enthusiasts.

The "Les Herbonautes" Website

The "Les Herbonautes" website was created to enable users to contribute to the digitization of the information on the plant labels. Users can choose from a range of missions that are selected from the Recolnat database (see Figure 1).

These missions feature pictures of plant discoveries, each requiring analysis of the accompanying labels. Once a mission is selected, participants are required to enter input for seven specific fields of information. These fields are country, region, date, collector, collection number, locality, and geographic coordinates (see Figure 2). To collect these data is the primary goal of the research. Thus, it is vital to ensure that each participant engages in a thorough and thoughtful analysis process.

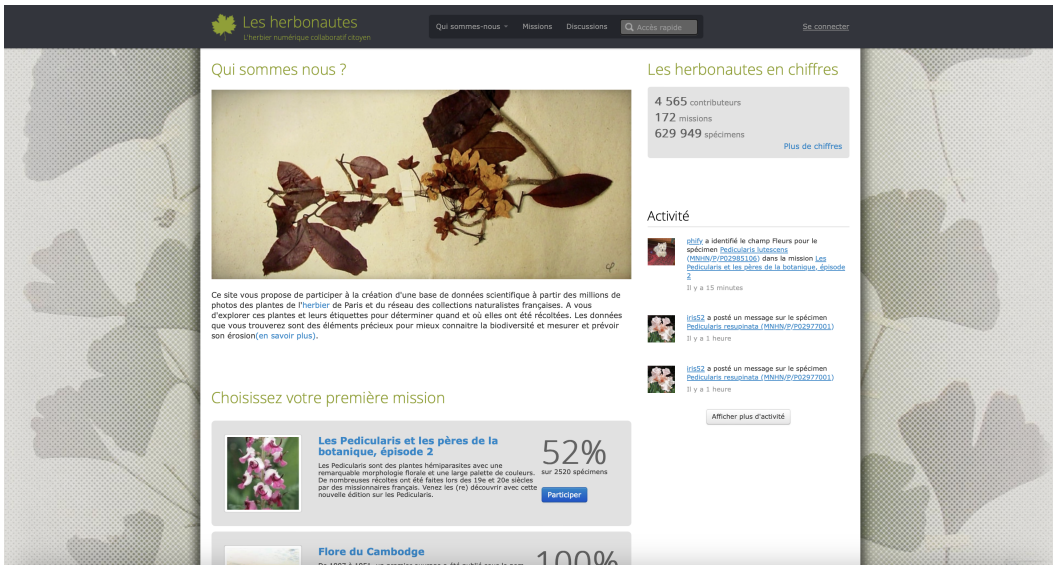


Figure 1. Homepage of Les Herbonautes.

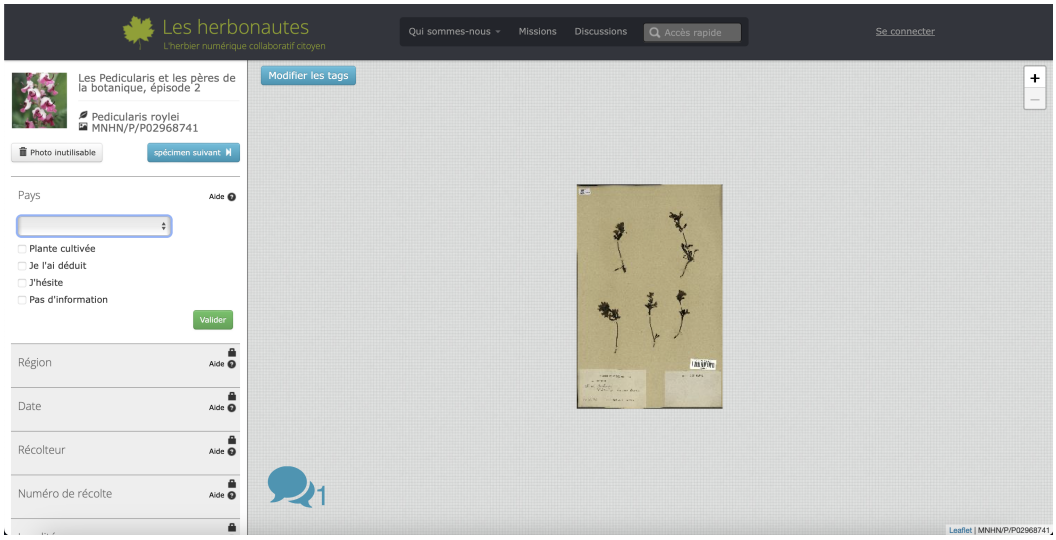


Figure 2. Mission-solving interface, where the left-hand side shows input fields.

In addition to data entry tasks, the website includes a group chat function. This feature allows participants to discuss their findings and share insights in real time with both fellow participants and admins. The group chat serves as a collaborative chat forum where users can ask questions, seek clarification, and receive feedback.

4. Problem Description

This section focuses on describing the specific requirements outlined by the representatives of the Muséum National d'Histoire Naturelle (MNHN) for the project "Les Herbonautes". It analyzes the context of use, considering the unique needs of its participants and the challenges they face. Finally, this section formalizes these requirements to provide a clear foundation for the subsequent development of the system.

4.1. Meeting with the Responsible Parties

During a first project meeting with representatives of the MNHN, who are responsible for the "Les Herbonautes" project, several key aspects of data quality, UX, and system efficiency were discussed. The focus was on identifying current challenges and exploring

potential solutions to enhance the overall functionality and user engagement of the platform. The main issues are discussed in this section.

4.1.1. Data Quality

One of the primary concerns about the quality of the collected data was the risk of erroneous data being validated on the platform. Because cross-validation is already implemented, this issue can only occur if multiple users accidentally enter the same incorrect information. But, if there are malicious intentions, there is a potential risk of deliberate abuse. A user could create multiple accounts to intentionally submit the same incorrect information, thus misleading the system and hindering the project's progress.

Another challenge discussed pertains to the entry of GPS coordinates for sample locations. Exact coordinates are often difficult for users to determine, especially since the precise GPS data are typically not included with the sample itself. Users must therefore deduce these coordinates, leading to inconsistencies. To mitigate this, the current solution involves calculating the mean of all submitted GPS coordinates. This solution might be problematic though, as there is no control on whether the mean of the coordinates is a better estimate for the actual plant discovery than any of the coordinates submissions alone.

To improve data quality, the system could utilize the information from completed missions more effectively. For example, if it is known that a particular collector has primarily gathered samples in Greenland, the system could prompt users when entering the name of that collector to consider Greenland as a country of discovery. Establishing connections between the region or country and specific collectors, especially when the collection dates of two samples are close to each other, could significantly enhance the data accuracy, decrease the time spent analyzing one sample, and reduce errors.

4.1.2. Quality of Life

The discussion also highlighted several quality-of-life improvements that could enhance UX on the platform. Currently, there is no automatic content control mechanism in the group chat, which leaves the platform vulnerable to both hurtful and incorrect content. Introducing a moderation system could help maintain a positive and constructive environment for all users while reducing the administrative resources needed. In the current state of the system, project administrators control the discussion in the group chat and participate in it when they detect incorrect information.

In general, the administrative workflow for uploading data to the platform could be improved. Administrators are required to download data from the Recolnat database, choose the samples that they want to include in one mission, manually create a CSV file out of them, and then upload the file to the "Les Herbonautes" platform. Automating this process would improve efficiency and reduce the workload of administrators, who already face limited resources.

A more engaging UI was also suggested to improve user engagement, as the current UI may be confusing for new users. Implementing a chatbot system could significantly improve UX by guiding them through the website, answering user questions, or suggesting the most relevant and interesting missions based on user preferences [19]. However, it was emphasized that the use of chatbots should remain optional to accommodate user preferences. The goal is to create a more engaging system, thus enhancing UX, which should increase user engagement and participation.

4.1.3. Efficiency

Efficiency improvements were another key focus of the system that should be implemented. An identified issue was that, if a sample picture lacks information for certain fields (e.g., the date of collection is not noted), it may be repeatedly shown to users even though

all existing information has already been extracted. This redundancy could be mitigated by developing a mechanism to recognize when a sample has reached its maximum data extraction potential and remove it from the rotation.

It was also noted that the chatbot could provide clear onboarding and guidance to users at all times, explaining the purpose of the project and what actions to take at any given time. This would help new users acclimate more quickly and reduce confusion [30].

Additionally, saving user preferences could allow the system to communicate differently depending on which user it interacts with, or to suggest missions tailored to individual interests, thus enhancing user engagement and efficiency [31].

4.2. User Analysis

The user base of the "Les Herbonautes" platform consists of a diverse group of participants who engage in the digitization and data collection efforts required for the project. However, there are several notable characteristics and trends that define the typical user profile on the platform.

Generally, participants in CS projects, including those on the "Les Herbonautes" platform, often share certain common traits. They are usually motivated by a personal interest in science and a desire to contribute to research efforts [32]. Many participants enjoy learning and expanding their knowledge in specific fields and find satisfaction in contributing to meaningful scientific endeavors. Furthermore, participants are often driven by a sense of community and collaboration, enjoying the opportunity to work with others who share their interests in nature and conservation [2]. The platform attracts a wide range of participants, many of whom may not have advanced technical skills or experience with digital tools. This variation in technological proficiency can impact the ease with which users interact with the platform and complete their assigned tasks. Ensuring that the platform is user friendly and accessible to those with limited technical skills is crucial to maintaining user engagement and participation rates.

CS platform users also exhibit different levels of expertise in biological research and data collection. Some users may have professional backgrounds in botany, ecology, or related fields, bringing a high level of expertise to the project. Others may be amateurs or enthusiasts with a keen interest in CS and biological research but without formal training [2]. This range of expertise presents both an opportunity and a challenge. While expert users can provide high-quality data and insights, less experienced users may require additional guidance and support to ensure the accuracy and reliability of their contributions.

Despite the openness of the "Les Herbonautes" platform to all skill levels, the majority of contributions are made by a relatively small group of expert users [2]. These core contributors are often highly motivated citizen scientists who have a deep interest in the project's goals and are committed to contributing regularly. This group is crucial to the platform's success, as their consistent participation and high-quality input drive much of the data collection process forward.

The limited number of new contributors on the "Les Herbonautes" platform might account for that. The influx of new users has been relatively low [2], which could be due to several factors, such as the platform's perceived complexity, the niche nature of the project, or the specific expertise required for meaningful participation. This scarcity of new participants poses a challenge to the platform's sustainability and growth, as it relies heavily on the continued efforts of its existing user base.

4.3. Context Analysis

The citizen scientists participating in the "Les Herbonautes" project typically operate within a specific context that influences their ability to contribute effectively to the project.

Understanding these contextual factors is essential for optimizing user engagement and ensuring the continued success of the project.

The nature of data collection and digitization tasks, such as transcribing handwritten labels and entering metadata, requires a device with a sufficiently large screen and reliable internet access. Although mobile devices could theoretically be used for some tasks, the precision and detail required for accurate data entry often make computers the preferred choice for participants.

Additionally, contributing to the "Les Herbonautes" project requires a significant investment of time. Participants must not only familiarize themselves with the platform and its protocols but also dedicate considerable time to carefully reviewing and entering data [2]. This often involves cross-referencing specimen information with external databases or reference materials. The large amount of time required for this implies that users actively plan to contribute to the project, rather than being able to work on it on the fly.

4.4. Formalized Requirements

Based on a meeting with the project representatives and the subsequent analysis, formal requirements for the system have been established, integrating the feedback provided by the project representatives. The requirements are listed in this section. To what extent these requirements can be fulfilled by the system implemented here will be discussed in the following sections.

4.4.1. Data Validation

R1: User Input Error Analysis: The system must analyze user input for errors and inconsistencies to ensure data accuracy. If the input is not validated, the user should be prompted to re-enter the data.

R2: Onboarding Process: The system must use project and context information to help new users familiarize themselves with the platform and data entry guidelines.

R3: User Preferences and Expertise Management: The system must save and utilize the preferences and expertise levels of each user to personalize their experience and thus enhance user engagement.

R4: Specimen Database Access: The system must have seamless access to the specimen database. It should support efficient data retrieval and entry for mission-related tasks.

4.4.2. Data Quality

R5: Input Recommendation: The system should recommend appropriate inputs for the fields.

R6: Required Fields Definition: The system should clearly define the required fields for data entry to maintain consistency and completeness across the database. Users should be guided to fill in all necessary fields accurately.

R7: Research Support: The system should assist users in researching specimen-related information by providing answers in natural language. This support will help users make informed decisions during data entry while lowering the barrier of participation for non-expert users.

R8: GPS Coordinate Handling: The system should manage GPS coordinates by automatically generating them after users have entered the information for all relevant fields. This process ensures the accuracy of the geographic data and decreases the amount of information users must deduce.

R9: Instant Feedback to Incorrect User Input: The system must provide instant feedback when users enter incorrect data. This immediate response allows users to correct their inputs promptly, enhancing the overall quality and reliability of the data collected.

4.4.3. User Experience

R10: Enhancement of Expert and User Communication: The system should improve communication between experts and citizen scientists. This can be achieved through a chatbot that allows users to contact experts.

R11: Admin Assistance: The system should provide tools and features to assist administrators in managing data and overseeing user contributions. This includes simplifying data uploads and moderating group chats.

R12: Mission Recommendation System: The system should recommend missions to users based on their preferences and prior activity on the project website. This feature aims to keep users engaged and motivated.

R13: Natural Language Interaction Interface: The system should utilize the NLP capabilities of LLMs to facilitate intuitive and conversational interactions between users and the system.

R14: Chat Moderation and Control: The system should moderate messages in the group chat to control inappropriate or incorrect content. The system should also be able to answer any user questions accurately and in a timely manner.

R15: Optionality of Chatbot Usage: The system should allow users to choose whether they want to solve missions using either the existing interface or the new chatbot.

R16: Clear System Identity: The system must clearly communicate to users that the chatbot system is an AI and not a human, maintaining transparency. This understanding helps to set appropriate expectations for user interactions.

R17: User-Friendly Interface: The system, and specifically the chatbot interface, must be understandable and easy to use for citizen scientists of all tech-savviness levels. The interface should facilitate smooth interaction and reduce user frustration.

R18: Abort Chat Functionality: The system must allow users to abort the chat at any time if they choose to disengage. This feature ensures a user-centered approach, respecting user autonomy and preferences.

5. Proposed Solution

To address the challenges mentioned in the prior sections and improve both data quality and UX, we propose creating a network of LLM agents and integrating it with a chatbot as an interface with the "Les Herbonautes" platform. LLM agents can serve multiple functions to enhance the project. First, they can provide real-time guidance and support to users, helping them understand the requirements of specific tasks and improving their ability to accurately transcribe information. By offering instant feedback and clarification, chatbots can help reduce errors early in the data entry process, decreasing the need for extensive post-entry validation.

LLM chatbots enable users to interact in natural language, significantly reducing the need for a deep understanding of technical systems or scientific jargon. This natural language interaction makes the platform more accessible to a broader audience, including those who may be less familiar with botanical terminology or the technical aspects of the digitization process.

In addition, LLM chatbots can serve as virtual assistants, guiding users through complex transcription tasks that would otherwise require direct human intervention. This not only democratizes access to expert knowledge but also fosters a more engaging and educational experience for users, further motivating them to contribute consistently and accurately.

Using NLP capabilities, chatbots could identify and flag potentially erroneous or inconsistent data entries in real time, prompting users to correct them before submission. This proactive approach to data validation can enhance the accuracy and reliability of the data being collected, ultimately contributing to a more efficient and scalable digitization effort.

By lowering these barriers to participation, LLM chatbots have the potential to increase user engagement and retention, attracting more diverse contributors to the project.

To more generally improve both UX and data quality with the project's focus on the data input workflow, the integration of the system should include features that seamlessly guide users back to the data collection process whenever necessary [33]. To prevent frustration during interactions, a simple mechanism for aborting the chat should be implemented, allowing users to exit the conversation at any time.

To address the varying levels of expertise among participants, the chatbot should be capable of categorizing users into different skill groups, offering tailored, complexity-appropriate responses, and missions. This approach helps users solve tasks more effectively and minimizes confusion and frustration [34].

Additionally, the system should proactively warn users about common errors during data entry to improve data quality.

For automated data processing, a separate Python script could analyze images and handwritten text, with the results reviewed by a human before being saved to ensure accuracy [33].

5.1. Conception

The conception of the new system for the "Les Herbonautes" platform focuses on enhancing UX and data quality through the integration of an interactive chatbot interface. Below, we detail the key design elements and considerations for the implementation of this system.

The most fundamental design feature is the chatbot interface. Citizen scientists will interact with the entire system through this interface. Allowing it to be movable across the screen provides users with the flexibility to position it according to their preferences and needs, ensuring that it does not obstruct important content or hinder their interaction with the platform, especially when analyzing pictures.

The visual design of the chatbot overlay will be consistent with the branding of the "Les Herbonautes" platform. To achieve this, the chatbot overlay will use the color palette and font design of Les Herbonautes, creating a cohesive look and feel that aligns with the platform's existing aesthetic.

To ensure transparency and manage user expectations, it is essential that the chatbot clearly communicates that it is not a human. This will be achieved through explicit statements within the chatbot's responses, as well as through design elements that distinguish it from human interaction. By making it clear that users are interacting with an automated system, we can set appropriate expectations for the type of assistance provided and reduce frustration.

The project will utilize open source LLMs to power the chatbot functionality. Open source models provide the opportunity to tailor the language processing capabilities to the specific needs of the "Les Herbonautes" platform, ensuring that the chatbot can handle the unique terminology and context of biodiversity research and CS without increasing the monetary resources necessary to host the project. The usage of open source LLMs guarantees that the privacy of users remains solely in the hands of the maintainers of the "Les Herbonautes" platform, avoiding data transfers to third parties.

A critical component of the system's architecture is the database that will store user information and mission details. This database is essential for personalizing the UX, as it will track user preferences, level of expertise, and past contributions. By storing this information, the system can provide customized recommendations for missions, offer relevant feedback, and facilitate more targeted interactions.

5.2. Proposed Architecture

The chatbot is the central decision-making component, determining the nature of each user query and routing it accordingly. Queries are categorized into three main types: onboarding, data entry, and questions. When first opening the chatbot overlay, users are greeted with a welcome message prompting them to interact with the chatbot.

5.2.1. Question Handling

When a user asks a question, the chatbot directs the query to the appropriate specialized LLM agent. If the query pertains to mission data, an SQL-based search retrieves the relevant information. For general questions that the system can answer without further data retrieval, a general-purpose answer LLM is contacted. If the query requires external information, the system performs a web search to provide an accurate response.

5.2.2. Onboarding

For onboarding, the system distinguishes between general and specific queries. General queries about the website's purpose and functionality trigger a general onboarding message. For more detailed questions, an LLM component should utilize additional information to answer the user's question.

5.2.3. Data Entry

The data entry process is the most intricate and critical part of the system, forming the central pipeline of the graph. Initially, users are asked if they prefer a mission suggestion based on their interests or if they want to select a mission themselves. If users opt for a suggestion, the system checks the database for their stored preferences. If such information does not exist, users are prompted to specify their interests, which are then saved for future use. An LLM analyzes these preferences and ranks the available missions. The user then selects a mission from the provided list, either based on their preferences or their own choice.

Once a mission is chosen, a random sample picture is selected as the current task. If the user has previously contributed to this sample, they are given the option to continue from where they left off. The system displays their progress, allowing them to confirm or adjust their previous entries. If they choose to start fresh or have no prior entries, the input process begins from the first field.

For each field, users are asked if they want the system to suggest input or enter it manually. If they request a suggestion, a Handwritten Text Recognition (HTR) algorithm processes the plant labels to extract the relevant information. An LLM matches this output to the available fields, and the suggested input is presented to the user for confirmation. Users can confirm or reject the suggested input, or they may enter their own data, which are then validated by another LLM. If the input is deemed invalid, the system prompts the user to provide a revised entry. Once validated, the input is saved, and the user moves on to the next field.

After completing all the required fields for a sample, the user is asked if they enjoyed the mission. If they did, the mission is saved to their interests for future recommendations. Users are then asked if they wish to continue with more missions. If they decline, the session ends with a thank you message. If they wish to continue, they can either proceed with a new sample from the same mission or start a new mission altogether, repeating the data entry process as described.

5.2.4. Group Chat Functionality

Whenever users enter messages in the group chat, an LLM agent scans the message and decides whether to respond to it or not. If the agent decides that the system should reply, the query is given to another LLM agent that decides what workflow to send it to: onboarding, question handling, or analyzing input. The question handling and onboarding workflows operate the same as described before except that the data entry workflow is not included, as this task should be realized with the user's private chatbot. When the user enters a harmful message according to the LLM agent, it is checked against a list of harmful words and analyzed by another LLM agent. If this results in a harmful content flag by the system, the message is deleted and the specific user informed by a private pop-up.

5.2.5. Admin Chatbot

Admins should have access to a chatbot that has more functionalities than the user chatbot. If they want to start a new mission, the chatbot asks them to give more details. Admins are prompted to input the name of the mission, given a quick description, and asked if they want to change the amount of redundant inputs for specific fields. Lastly, they input the image files and the system creates a new mission in the database of the project, which is then synchronized with the website.

5.3. Discussion of the Proposed Architecture

The proposed architecture of the chatbot system is structured into different workflows to meet the varying requirements and efficiently partition the system's functionalities. This modular approach allows for a clear separation of concerns, ensuring that each workflow can be optimized independently for its specific purpose. By adopting this architecture, the system effectively meets the primary requirements, facilitating robust and scalable performance.

6. Implementation

This chapter outlines the implementation of the enhanced system for the "Les Herbonautes" platform. It provides a detailed explanation of the chosen framework, LangGraph, and describes the specific architecture developed to integrate the chatbot system.

6.1. Challenges

The implementation of the system faced several challenges, particularly related to the use of open source LLMs. One significant issue was that open source LLMs are currently not fully compatible with tool usage in the LangGraph framework. This limitation restricted the functionality of the chatbot system, as certain tools and features within LangGraph could not be utilized effectively with open source models.

Another challenge with open source LLMs is the potential for flawed decision making. Unlike more advanced, proprietary models, open source LLMs may lack the sophisticated decision-making capabilities needed to accurately determine the next steps in a conversation or interaction. This limitation could lead to suboptimal or incorrect responses, affecting the overall workflow.

Additionally, the open source LLMs imposed constraints on the length of prompts that could be used. The inability to use longer prompts restricted the complexity and depth of queries and responses that the system could handle, limiting the chatbot's ability to provide comprehensive answers or engage in more detailed interactions with users.

Although the initial plan included integrating a chatbot capable of moderating and participating in group discussions, the limited timeframe necessitated prioritizing other core functionalities. For the same reason, the intended functionalities for admins were not

implemented (an admin-specific chatbot, users being able to contact admins through the network of agents). Additionally, the implementation of a clever subsystem that chooses the most accurate GPS coordinates from the different user inputs and information about the region of a herbarium sample proved to be a complex problem of its own. Since no solution could be found that reliably improved the GPS coordinates, the old mechanism was kept (see Section 4).

6.2. Architecture

The implementation of the enhanced system closely followed the initial conception, adapting the proposed workflows to fit within the graph structure provided by the LangGraph framework. The system can be described as a state graph as shown in Figure 3. Here, the different workflows are visible, which were tailored to meet the specific needs of the users. Purple components encode python scripts, while every blue component is an LLM agent based on a specific LLM. Any gray components are simple chat messages that are always sent to the user if their node is reached. The orange diamonds are points where the system expects user input.

The LangGraph framework (https://www.langchain.com/langgraph, accessed on 27 October 2024) was used to create a network of interconnected LLM agents.

LangGraph is a versatile framework designed to facilitate the creation and management of graph-based structures that integrate multiple language models for complex conversational tasks. The framework is particularly well suited to applications that require sophisticated interaction patterns, data processing, and decision-making capabilities, such as those found in CS projects such as "Les Herbonautes". LangGraph enables the development of a network of interconnected LLM agents, each with specific roles and responsibilities. This graph-based architecture allows for efficient task delegation and parallel processing of user queries, making it ideal for environments where diverse types of interactions and data handling are required.

Ollama (https://ollama.com, accessed on 13 December 2024), that was used as a framework for building and running LLMs locally, is a platform designed to simplify the deployment and management of LLMs on local machines. It provides an efficient environment for running open source LLMs without requiring extensive cloud infrastructure. By utilizing Ollama, the implementation could leverage the power of open source LLMs while maintaining control over the data and reducing dependency on external servers, albeit with some trade-offs in processing speed and model capabilities.

The open source LLMs utilized in this project were Llama 3 and SQL-coder. SQL-coder was used only for the specific task of creating MySQL queries to interact with the project's databases, for which it was trained. The other agents on the network used Llama 3 without additional training for any NLP tasks. Since the same queries should produce the same replies and agents should precisely follow the instructions given by their respective prompts, the temperature for every LLM agent was set to 0. The temperature of an LLM is a measure of the randomness in its output generation, controlling how deterministic or creative its responses are. Lower temperatures make the output more focused and predictable, while higher temperatures encourage more varied and imaginative responses. Apart from how the agents were interconnected, the prompts for them were the most important customization tools. Changing only small parts of a prompt greatly influences an LLM's output. Each prompt had to be individually created and carefully improved by testing.

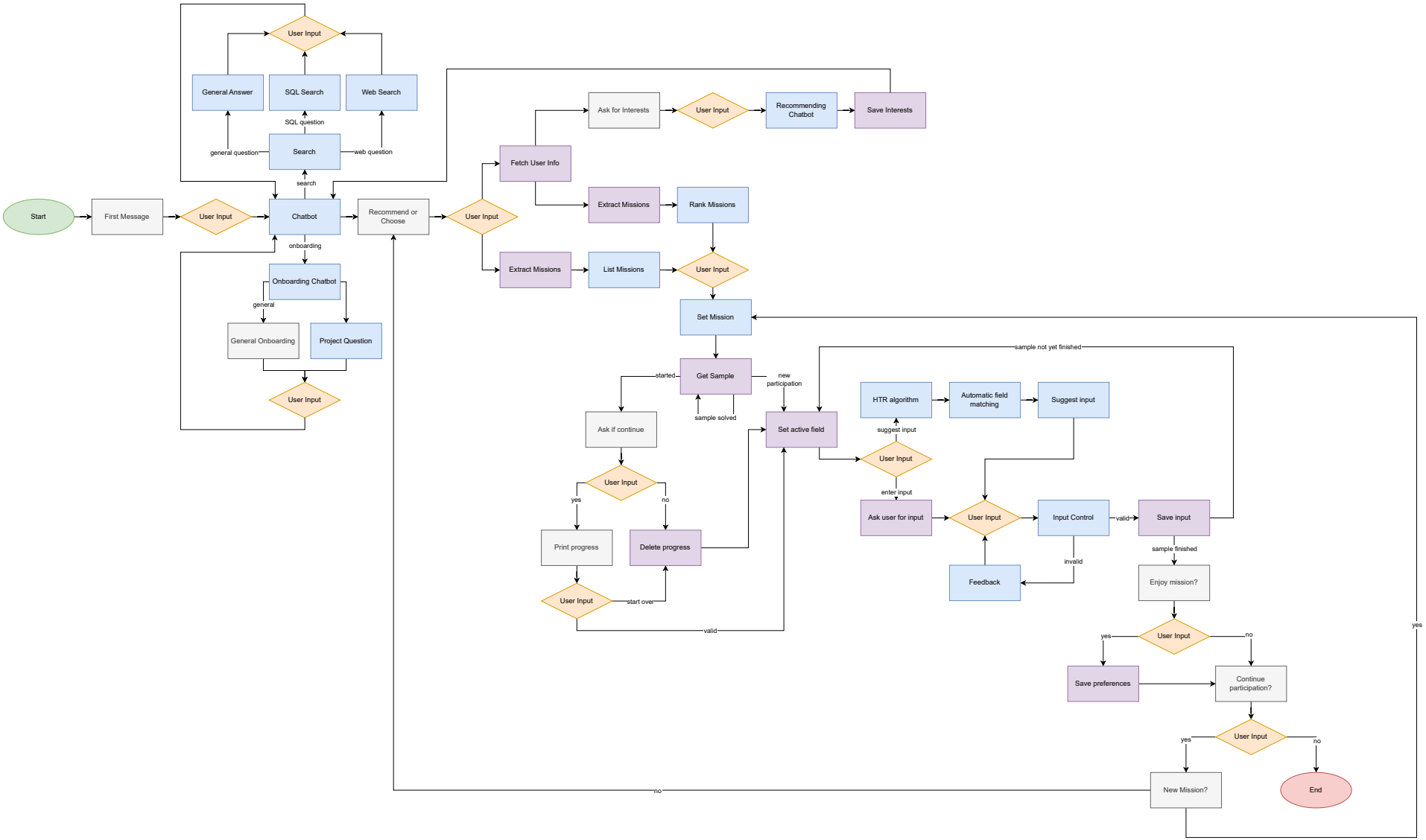


Figure 3. State graph of the realized network of LLM agents for the "Les Herbonautes" chatbot.

Instead of relying on the tool functionality of LangGraph, the system was restructured to utilize the nodes within the LangGraph framework. Conditional edges were established from the LLM agents to these nodes, allowing dynamic decision making based on user input and contextual data. This approach provided greater flexibility in managing the conversation flow and enabled more precise control over the system's responses.

For the database requirements, MySQL (https://www.mysql.com/ accessed on 27 October 2024) was chosen to host the necessary databases. MySQL offered a reliable and scalable solution for storing user information, mission data, and other critical information required for the system to function effectively. This choice ensured that data could be accessed and updated efficiently, supporting the system's need for real-time data management and interaction.

For each query, an LLM agent decides which particular subgraph of the system to send it to. There are four subgraphs implemented as separate workflows: data input, mission recommendation, onboarding, and question answering.

6.2.1. Onboarding Workflow

New users can ask the chatbot for guidance on how to navigate the platform, participate in missions, and contribute effectively to the project. An LLM agent (Onboarding Chatbot, see Figure 3) will then categorize the query into general onboarding (see Figure 3), in which case a general answer in the form of a predefined chat message will be given to the user that briefly describes the project's goals and the user's purpose on the website, or project question, where another agent uses Retrieval-Augmented Generation (RAG) [35] on documents prepared for onboarding to provide more detailed answers.

6.2.2. Mission Recommending Workflow

Another key workflow facilitated by the proposed system is the mission recommendation. Users can ask the chatbot to recommend a mission based on their interests and past activities on the platform. In this case, the system extracts user preferences from the user database, and information on the active missions, which are then analyzed by an LLM agent (rank missions, see Figure 3) to derive the most relevant missions for the specific user. The chatbot then presents the ranked missions to the user. If a user requires missions to be recommended to them, but no prior information on the user has been saved, the system asks the user for their interests (Ask for Interests, see Figure 3). The user can then decide whether they want to enter information on their preferences or choose a mission themselves.

6.2.3. Question Answering Workflow

When a user asks a question, it is sent to the question answering pipeline. An LLM agent (Search, see Figure 3) categorizes the query into general questions that the LLM agent could answer without additional information, specific questions that require a web search to answer them, or data-related questions for which the specimen database should be searched. For the first category, an LLM agent (General Answer, see Figure 3) simply answers the question based on its pre-trained knowledge. If the user query belongs to the second category, a different LLM agent (Web Search, see Figure 3) uses the Tavily search engine tool (https://tavily.com, accessed on 13 December 2024) for a web search to find more information on the subject of the question. Lastly, if the system needs specific mission- or specimen-related information to answer the user question, an agent that uses SQL-coder as LLM (SQL Search, see Figure 3) creates a MySQL query to extract the relevant data.

6.2.4. Data Input Workflow

The data input workflow is a core functionality of the system, enabling users to enter information for specific fields from pictures of plant samples. Users can input data such as country, region, date, and other relevant details directly through the chatbot interface. This process follows the same structure as the existing system's (see Section 2) in order to change the actual workflow as little as possible. This allows users to quickly adapt to the new system. With the proposed solution, users could also ask the chatbot for assistance if they have questions about the data entry process or require clarification on specific fields. In addition, users can request input recommendations from the chatbot. In this case, the network of LLM agents leverages OCR analysis and NLP capabilities to suggest the most likely entries for each field, which the chatbot passes on to the user.

When a user starts the mission-solving process, their expertise level is controlled to check which fields they are allowed to enter information for. Additionally, only data on nonvalidated fields are requested, so the system scans the database to check which field has less than at least two identical inputs from different users (Set active field, see Figure 3).

If the user asks for an input recommendation, an OCR algorithm scans the specific image for legible information. The results are given to an LLM agent that tries to match them to the seven fields of relevant information (see Section 3 for more information). When data are found for the field the user is working on, the chatbot passes them on to the user to control them. Once the user submits their input, an LLM agent (Input Control, see Figure 3) immediately reviews it. This agent is responsible for semantically validating the entered data to ensure accuracy and consistency. So, if the user enters a region that is not inside the country they entered before that, the agent will inform the system that the input is invalid and give a reason for its decision. Likewise, if the user enters a collection date that is in the future, the agent will deem the input invalid and give appropriate feedback. If the input is deemed invalid or contains errors, another LLM agent (Feedback, see Figure 3) creates feedback for the user based on the prior checks, highlighting the issues and prompting them to correct their entries, which the chatbot then sends to the user. If the input is decided to be valid, it is saved (Save Input, see Figure 3) and the system moves to the next field, unless the user does not have the required expertise level, or every field for this specific sample has been validated. Once a user finishes entering information for a sample, the system inquires whether they enjoyed the mission (Enjoy Mission, see Figure 3). If they did, it is added to their interests and used for future mission recommendation (save preferences, see Figure 3).

Figure 4 shows an excerpt of an interaction with the implemented system that users can interact with by interacting with a chatbot interface. The excerpt is used as an example of the functionalities that were explained.

CHATBOT:
Welcome to Les Herbonautes! I am here to assist you with any questions or problems you might have. Would you like to work on a mission, or do you have a specific question? Let me know if you need assistance!

USER:
What is this project about?

CHATBOT:
This project is about digitizing the information from plant labels that have so far been recorded only analogeous. Users, like you, can help us by trying to read the handwritten plant labels and extract information from them. Do you have any questions about this, or do you want to start with a mission? You could ask me to suggest missions to you based on your interests, or choose a mission yourself. Just let me know if I can help!

USER:
I want to solve a mission.

CHATBOT:
Would you like me to suggest suitable missions to you, or would you rather choose a mission yourself out of the list of available ones?

USER:
Suggest a mission to me.

CHATBOT:
Based on your interests in roses, trees, and France, I've ranked the available missions as follows: 1. Rosa: Since you have a strong interest in roses, this mission is a natural fit, allowing you to explore and learn more about these beautiful flowers. 2. Cunoniacee: As you're interested in trees, this mission focuses on a family of flowering plants that includes many tree species, making it a great match for your curiosity about arboreal life. 3. Taraxacum: Although not directly related to roses or trees, the fact that you have an interest in France and this mission is often associated with French culture (dandelions are also known as "pissenlit" in French), makes it a suitable choice for you. These three missions align well with your interests, and I hope you find them engaging and informative!

USER:
I want to work on the Rosa mission.

CHATBOT:
Please enter information you can find on this label for which country the specimen was found in. We are looking for current countries, so if the mentioned country does not exist anymore today, please try and find a current country corresponding to the old one and let me know that you guessed or deduced it. I can also analyse the written information on the sample and suggest my interpretation to you. Ask me any questions you have about this! If you are unsure about the entered information, please also mention it. If you cannot find any information for the country, you can also tell me.

USER:
I have a question: In which country did Carl Ludwig Ritter von Blume work?

CHATBOT:
Here is a summary of the search results: Carl Ludwig Blume (1796-1862) was a German-Dutch botanist who worked primarily in the Netherlands. He was born in Braunschweig, Germany, but spent most of his professional life as Director of the Rijksherbarium at Leiden University. The standard abbreviation "Blume" is used to indicate him as the author when citing botanical names.

USER:
Now suggest input to me please.

CHATBOT:
Do you think this information could be correct? *Country*: Spain

USER:
Yes.

CHATBOT:
Thank you for helping to digitize information! All the information for this specific sample have been entered and saved to our database. Did you enjoy this particular mission?

Figure 4. Chat example from Les Herbonautes chatbot, containing onboarding, preference saving, mission solving, input suggesting and question answering aspects.

7. Evaluation

After implementing the system, an evaluation was carried out to determine whether it effectively improved the "Les Herbonautes" project. This chapter details the methods used

in the evaluation, the participants, and the results, providing an analysis of the system's effectiveness in improving the UX and data quality of the "Les Herbonautes" project.

7.1. Goal

The evaluation objectives were threefold: first, to determine whether the introduction of the chatbot system positively influenced UX on the Les Herbonautes platform; second, to assess whether the quality of data collected through the platform improved as a result of the chatbot's usage; and third, to explore any potential correlation between improved UX and enhanced data quality. Given the limited number of participants, our evaluation strategy relied mainly on qualitative feedback from experts, focusing on the functionality and potential improvements of the chatbot system, as quantitative measures were not conclusive. While other UX evaluations (e.g., broader user groups) could complement our evaluation, the expert-based approach was considered more feasible and valuable given the scope and constraints of our work, offering the advantage of leveraging specialized knowledge to evaluate complex system features.

7.2. Methodology

For the evaluation of the chatbot system, a separate website had to be implemented to closely resemble the original "Les Herbonautes" platform, particularly in terms of its functionalities. This mockup aimed to replicate the essential features of the "Les Herbonautes" website to provide a realistic environment to test the new chatbot system. To do so, only the Homepage and Missions workflows were implemented. The museum supplied datasets for five different missions. In each mission, participants could only enter information for the first three fields (country, region, and collection date).

Although all the functionality of the original website and the entire range of initial chatbot features (see Section 6.1 for details) could not be fully implemented due to time constraints, the mockup provided a focused environment to test the core capabilities of the system. This methodology allowed for an efficient evaluation within the available timeframe, ensuring a meaningful assessment of the system's performance and interaction features.

The mockup website, including the chatbot system, was hosted on a virtual machine (VM) provided by the University of Lübeck. The absence of a GPU on this VM meant that the locally run LLMs used by the chatbot took longer to process user queries.

Participants were asked to solve three different missions using the mockup platform to simulate typical user interactions. They were told to enter input for the first three fields of information (country, region, and collection date). After completing the missions, they were invited to complete the System Usability Scale (SUS) and User Experience Questionnaire (UEQ) to quantitatively measure usability and UX. Additionally, a semi-structured interview was conducted to gather qualitative insights into the participants' experiences, providing a more comprehensive understanding of the chatbot system's impact on both user satisfaction and data quality.

Participants

The evaluation involved four participants from the MNHN, all of whom were experts on the "Les Herbonautes" system. We decided to involve experts, because they were familiar with the system they were evaluating and had knowledge of the relevant information and workflows of the specific use case across all user levels. Thus, their feedback naturally anticipated insights from other types of UX, providing valuable input for the evaluation. Furthermore, an evaluation with experts requires significantly fewer resources while achieving valuable qualitative results. The participants were all females involved in the "Les Herbonautes" project, had a scientific background, and their mean age was 50.25 years.

7.3. Results

Qualitative feedback from the semi-structured interviews provided deeper insight into the functionality and areas for potential improvement of the chatbot system. For example, participant 1 suggested that the system should include references to its research sources to improve trust and credibility. Also, participants 1 and 2 mentioned that, without sources or explanations for the system's answers, they were unsure as to how far they would trust the results of the system. In general, though, when asked if the participants trusted the system's answers, they answered that they did. The system was able to recognize all intentional wrong input from experts and provided feedback accordingly. All participants stated that they were aware that the chatbot was a computer system at any time. Two participants mentioned that they enjoyed interacting with the chatbot. All participants said that the interface was easily understandable and unobtrusive during mission solving. Three participants noted that the long calculation times negatively affected the overall experience. There were also errors when LLM agents selected the incorrect continuation node, leading to inaccurate responses or misdirected queries. This happened to one participant twice. Despite these challenges, all participants were generally impressed with the range of questions the system could answer accurately, highlighting its potential as a valuable tool for user engagement and support on the "Les Herbonautes" platform.

Table 2 shows the results of both the UEQ and the SUS. As the UEQ has a range of 1–7 and the range of the SUS is 1–5, the results of 4.8 and 3.125, respectively, show that the system is perceived as moderately useful, usable, and effective. These rather negative results could be due to the errors also mentioned in the interviews. Since system errors heavily affect user trust in a system [36], and trust is an important factor in user engagement in a context where the user has to rely on system predictions [37], even small system errors or long calculation times can lead to more negative reviews of the system. Because the system is a prototype, future system improvements could enhance the ratings of the questionnaires. Nevertheless, LLMs are non-deterministic, and thus the right choice of continuation node might never be completely accurate. The expertise of the participants might have also decreased the rating: as all experts are proficient in data input with the existing system, interacting and getting used to the new system required more resources in the form of time and energy, which decreased perceived UX. However, these results are not statistically significant due to the limited number of participants.

Table 2. Results from UEQ and SUS.

Participant ID	UEQ	SUS
1	5.4	3.3
2	4.1	2.7
3	4.9	3.1
4	4.8	3.4
Average	4.8	3.125

The results of the analysis of the data that were entered during mission solving contained no mistakes compared to the validation labels. In total, data for 36 fields were entered and every piece of user input was correct. This could be an indication of improved data quality due to additional data validation strategies, but the number of results is not enough to imply statistical significance, and the participants were experts in the data collection process, so their performance is not representative.

In addition to the expert evaluation, the original list of requirements (see Section 4) was reviewed and completion assessed. Of the 18 requirements initially defined, 13 were successfully fulfilled. The following sections highlight key requirements, describing the

implementation approach to fulfill them or reasons why they have not been implemented. Five requirements—R4: Specimen Database Access, R8: GPS Coordinate Handling, R10: Expert and User Communication Enhancement, R11: Admin Assistance, and R14: Chat Moderation and Control — remain unimplemented due to project constraints and specific external limitations that prevented full integration.

R1: User Input Error Analysis: The system must analyze user input for errors and inconsistencies to ensure data accuracy.

An automated mechanism for analyzing user input was implemented to detect errors and inconsistencies. An LLM agent flags potential errors and offers corrective feedback to users in real time based on the user input and the field for which to find information, thus improving the overall quality of the collected data.

R4: Specimen Database Access: The system should have seamless access to the specimen database.

Although the aim was to achieve seamless access to the project sample database, restricted access to this database was a limiting factor. Consequently, an alternative, independent database was used for system testing and evaluation. The system was built to be database-agnostic, meaning that it can theoretically connect to any compatible database, as long as the IP address is provided. This flexibility allows future integration with the actual project database if permissions and access are granted. However, without direct access to the project database, verification of this functionality within the intended environment was not possible.

R5: Input Recommendation: The system should recommend appropriate inputs for the fields.

The system uses Optical Character Recognition (OCR) algorithms to facilitate data entry by providing intelligent input recommendations. OCR technology analyzes specimen labels or other relevant documents, extracting and interpreting text to populate specific fields within the system. Following this, an LLM agent controls and matches the extracted text to the relevant input fields and suggests OCR information that could be correct to users asking for suggestions. This design allows users to review the recommended input and verify its accuracy, ultimately increasing data quality and reducing the resources needed by the users.

R7: Research Support: The system should assist users in researching specimen-related information by providing answers in natural language.

The system incorporates features to help users research specimen-related information by responding to questions in natural language. This is achieved through a multifaceted approach that includes a web search, general project data retrieval, and specific mission data queries. The system can dynamically generate SQL queries to retrieve relevant data from previously collected mission data. Additional web searches provide comprehensive responses enriched by project-specific knowledge. This approach empowers users to gain deeper insights into specimen information, reducing the research time needed.

R8: GPS Coordinate Handling: Implement a mechanism for managing GPS coordinates by automatically generating them.

The goal was to implement an automated GPS handling mechanism for capturing and managing geographic coordinates to improve the accuracy of locality information. Due to the technical complexity of implementing accurate GPS functionality and integrating it with mapping services, this requirement was not fulfilled. Precise handling and validation of GPS data require specialized software tools and extensive testing to ensure accuracy.

R9: Instant Feedback to Incorrect User Input: The system must provide instant feedback when users enter incorrect data.

This feedback mechanism cross-references user input with predefined criteria, allowing the system to flag inaccuracies immediately. By informing users of errors at the point of entry, the system promotes more accurate data entry, reduces correction time, and increases data quality. The instant feedback loop allows users to address errors in real time, improving the overall UX and the reliability of submitted data.

R10 (Enhanced Communication), R11 (Admin Assistance), and R14 (Chat Moderation).

Although R10, R11, and R14 were initially outlined as requirements, external limitations prevented their implementation. Specifically, the integration of new communication or moderation tools would require additional access to the project's underlying codebase and careful consideration of data privacy. Consequently, these functionalities were not fully realized in the current version of the system. However, they remain viable candidates for future iterations, pending appropriate access and resource allocation.

R16: Clear System Identity: Clearly communicate to users that the chatbot system is an AI and not a human, maintaining transparency.

The chatbot clearly communicates its identity as an AI-driven assistant by openly introducing itself as an AI in user conversations and providing status updates during interactions, such as "Answer is being calculated". These notifications reinforce transparency, helping users understand that they are engaging with an automated system rather than a human and thus keeping user expectations realistic.

R17: User-Friendly Interface: Design the chatbot interface to be understandable and easy to use for users of all tech-savviness levels.

The system interface was designed with accessibility and ease of use as central priorities. The chatbot offers language support in three languages, and the overlay interface is freely movable and closable, enabling users to tailor its position and visibility to their preferences.

Through the successful fulfillment of 13 of 18 key requirements, this system provides a solid foundation for data accuracy, usability, and research support. Although five requirements were deferred, these core features establish a robust system, addressing the primary project goals and providing the planned functionality to users.

8. Conclusions

This paper explored the integration of a network of open source LLM agents into the Les Herbonautes platform as an innovative approach to enhancing UX and, consequently, data quality in CS projects. The evaluation conducted with experts from the Muséum National d'Histoire Naturelle revealed that the chatbot system was a promising tool, providing useful assistance and guidance during data entry and mission-solving tasks. These findings suggest that the use of LLMs in CS projects is a promising new area of research. By automating routine tasks and providing real-time feedback, LLM agents have the potential to reduce the burden on human resources and improve the overall efficiency of data collection efforts. This approach is particularly valuable in scenarios where expert guidance is limited.

However, the study also highlighted some challenges associated with the use of LLMs, particularly open source models. One of the main issues is their non-deterministic nature, which can lead to inconsistent outputs and decision-making errors. This limitation underscores the need for further refinement and development of these models to improve their reliability and accuracy in critical applications.

The integration of automatic handwriting recognition opens a unique opportunity to enhance the possible applications of networks of LLM agents in CS projects further, especially in contexts where users need to transcribe handwritten specimen labels. This feature can reduce user errors and unify the data entry process.

Overall, this research highlights the potential benefits and challenges of integrating LLMs into CS platforms. By combining multiple open source LLMs in a graph-like architecture and attempting to solve different, complex tasks related to the data collection process in CS projects, this research showed the potential benefits of utilizing LLM agents for data validation, question answering, onboarding, and recommendation tasks. This network of agents enables the processing of complex tasks that priorly needed human resources to accomplish, thus saving time and monetary resources. Additionally, it has a chatbot as an interface for user interaction, enabling conversations in natural language. Similar systems that utilize task division can be designed to fit any use case, though they perform better in surroundings where tasks can be divided into small and clearly defined subtasks. More evaluations should be conducted to prove the effectiveness of the implemented system.

9. Future Work

One important avenue for future work is the integration of additional user groups into the chatbot system, particularly by developing a chatbot for administrators. An admin chatbot could assist administrators during routine tasks such as managing data entries and generating reports from the database. By providing automated support to administrators, the system could reduce their workload and improve overall efficiency, allowing them to focus on tasks that require human judgment and expertise.

Another promising direction is the development of the group chat chatbot. Although the current system includes basic functionality to interact with individual users, enhancing the capabilities of a chatbot in a group chat setting could promote better communication and collaboration among participants.

Combining the features of LLMs and knowledge graphs [38] is also interesting for use within chatbots. In our future work, we will investigate constructing knowledge graphs from users' input with the help of LLMs [39] for further making use of the information provided by users in CS projects.

Exploring the use of commercial LLMs (proprietary LLMs with advanced capabilities) could also be a valuable area of future research. While the current implementation relies on open source LLMs, commercial LLMs may offer superior performance, especially in terms of decision making, natural language understanding, and response generation.

Furthermore, future improvements might include adding sources to the research results generated by the system. Providing users with citations or references for the information provided by the system would improve its credibility and trustworthiness. This feature is particularly important in a scientific context, where accuracy and verifiability are essential.

Author Contributions: The authors contributed to this paper as following: Conceptualization, A.-L.K., S.S. and S.G.; methodology, A.-L.K., S.S., S.G., J.G. and H.K.; software, A.-L.K.; validation, S.S., S.G., J.G., H.K., R.V.-L., M.P. and E.P.P.; formal analysis, A.-L.K.; investigation, A.-L.K., S.S., S.G., J.G. and H.K.; resources, R.V.-L., M.P. and E.P.P.; data curation, A.-L.K., S.S., R.V.-L., M.P. and E.P.P.; writing—original draft preparation, A.-L.K., S.S. and S.G.; writing—review and editing, A.-L.K., S.S., S.G., J.G., H.K., R.V.-L., M.P. and E.P.P.; visualization, A.-L.K.; supervision, S.S. and S.G.; project administration, S.S. and S.G.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the France 2030 investment program (ANR-20-SFRI-0013).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CS	Citizen Science
LLM	Large Language Model
UX	User Experience
CA	Conversational Agent
NLP	Natural Language Processing
NLG	Natural Language Generation
MNHN	Muséum Natural d'Histoire Naturelle
UI	User Interface

References

1. Bonn, A.; Richter, A.; Vohland, K. *Grünbuch Citizen Science Strategie 2020 für Deutschland*; GEWISS: Berlin, Germany, 2020.
2. Rouhan, G.; Chagnoux, S.; Dennetière, B.; Schäfer, V.; Pignal, M. The herbonauts website: Recruiting the general public to acquire the data from herbarium labels. In Proceedings of the Botanists of the Twenty First Century: Roles, Challenges and Opportunities, UNESCO International Conference, Paris, France, 22–25 September 2014.
3. James, T. *Improving Wildlife Data Quality: Guidance on Data Verification, Validation and Their Application in Biological Recording*; National Biodiversity Network: London, UK, 2006; Volume Guidance Manual.
4. Commission, E.; Centre, J.R.; Mitton, I.; Tricarico, E.; Schade, S.; Lopez Canizares, C.; Tsiamis, K.; Gervasini, E.; Adriaens, T.; Cardoso, A.; et al. *Data-Validation Solutions for Citizen Science Data on Invasive Alien Species*; Publications Office of the European Union: Luxembourg, 2021. [CrossRef]
5. Stein, C.; Teubner, T.; Morana, S. Designing a conversational agent for supporting data exploration in citizen science. *Electron. Mark.* **2024**, *34*, 23. [CrossRef]
6. Junior, S.B.; Ceravolo, P.; Groppe, S.; Jarrar, M.; Maghool, S.; Sèdes, F.; Sahri, S.; Keulen, M.V. Are Large Language Models the New Interface for Data Pipelines? In Proceedings of the International Workshop on Big Data in Emergent Distributed Environments, Santiago, Chile, 9–15 June 2024.
7. Weaver, W.N.; Ruhfel, B.R.; Lough, K.J.; Smith, S.A. Herbarium specimen label transcription reimagined with large language models: Capabilities, productivity, and risks. *Am. J. Bot.* **2023**, *110*, e16256. [CrossRef]
8. Ahmed, S. An Architecture for Dynamic Conversational Agents for Citizen Participation and Ideation. Ph.D. Thesis, Technische Universität München, München, Germany, 2019.
9. Bittner, E.; Oeste-Reiß, S.; Leimeister, J.M. Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. In Proceedings of the 52nd Hawaii International Conferenceon System Sciences, Maui, HI, USA, 8–11 January 2019. [CrossRef]
10. Dam, S.K.; Hong, C.S.; Qiao, Y.; Zhang, C. A Complete Survey on LLM-based AI Chatbots. arXiv **2024**, arXiv:2406.16937. [CrossRef]
11. Kim, S.; Lee, J.; Gweon, G. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In Proceedings of the CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow Scotland, UK, 4–9 May 2019; ACM: New York, NY, USA, 2019; pp. 1–12. [CrossRef]
12. Binns, R.; Van Kleek, M.; Veale, M.; Lyngs, U.; Zhao, J.; Shadbolt, N. 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Montreal, QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 1–14. [CrossRef]
13. Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21, Virtual, 3–10 March 2021; ACM: New York, NY, USA, 2021; pp. 610–623. [CrossRef]
14. Kvale, K.; Freddi, E.; Hodnebrog, S.; Sell, O.; Følstad, A. Understanding the User Experience of Customer Service Chatbots: What Can We Learn from Customer Satisfaction Surveys? In *Chatbot Research and Design*; Springer: Cham, Switzerland, 2021; pp. 205–218. [CrossRef]
15. Vadapalli, J.; Gupta, S.; Karki, B.; Tsai, C.H. Incorporating Citizen-Generated Data into Large Language Models. In Proceedings of the 25th Annual International Conference on Digital Government Research, dg.o 2024, Taipei, Taiwan, 11–14 June 2024; ACM: New York, NY, USA, 2024; pp. 1023–1025. [CrossRef]
16. Lee, G.; Hartmann, V.; Park, J.; Papailiopoulos, D.; Lee, K. Prompted LLMs as Chatbot Modules for Long Open-domain Conversation. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023, Toronto, ON, Canada, 9–14 July 2023; pp. 4536–4554. [CrossRef]

17. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 1877–1901.
18. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.H.; Le, Q.V.; Zhou, D. Chain-of-thought prompting elicits reasoning in large language models. In Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22, New Orleans, LA, USA, 28 November–9 December 2022; Curran Associates Inc.: Red Hook, NY, USA, 2024.
19. Radziwill, N.M.; Benton, M.C. Evaluating Quality of Chatbots and Intelligent Conversational Agents. *arXiv* **2017**, arXiv:1704.04579. [CrossRef]
20. van der Goot, M.J.; Hafkamp, L.; Dankfort, Z. Customer Service Chatbots: A Qualitative Interview Study into the Communication Journey of Customers. In *Chatbot Research and Design*; Følstad, A., Araujo, T., Papadopoulos, S., Law, E.L.C., Luger, E., Goodwin, M., Brandtzaeg, P.B., Eds.; Springer: Cham, Switzerland, 2021; pp. 190–204.
21. Klopfenstein, L.; Delpriori, S.; Malatini, S.; Bogliolo, A. The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms. In Proceedings of the DIS '17: Designing Interactive Systems Conference 2017, Edinburgh, UK, 10–14 June 2017; pp. 555–565. [CrossRef]
22. Urzedo, D.; Sworna, Z.T.; Hoskins, A.J.; Robinson, C.J. AI chatbots contribute to global conservation injustices. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 204. [CrossRef]
23. Sánchez Cuadrado, J.; Pérez-Soler, S.; Guerra, E.; De Lara, J. Automating the Development of Task-oriented LLM-based Chatbots. In Proceedings of the ACM Conversational User Interfaces 2024, CUI '24, Luxembourg, 8–10 July 2024; ACM: New York, NY, USA, 2024; pp. 1–10. [CrossRef]
24. Elliott, M.; Fortes, J. Using ChatGPT with Confidence for Biodiversity-Related Information Tasks. *Biodivers. Inf. Sci. Stand.* **2023**, *7*, e112926. [CrossRef]
25. Elliott, M.J.; Fortes, J.A.B. Toward Reliable Biodiversity Information Extraction From Large Language Models. In Proceedings of the 2024 IEEE 20th International Conference on e-Science (e-Science), Osaka, Japan, 16–20 September 2024; IEEE: New York, NY, USA, 2024; pp. 1–10. [CrossRef]
26. Mora-Cross, M.; Calderon-Ramirez, S. Uncertainty Estimation in Large Language Models to Support Biodiversity Conservation. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track), Mexico City, Mexico, 16–21 June 2024; pp. 368–378. [CrossRef]
27. Elliott, M.; Luciano, M.; Fortes, J. Integrating Large Language Models and the iDigBio Portal for Conversational Data Exploration and Retrieval. *Biodivers. Inf. Sci. Stand.* **2024**, *8*, e142696. [CrossRef]
28. Du, Y.; Wang, Y.; Zhao, E. Leveraging Multimodal LLMs for Plant Species Identification and Educational Insights. *SocArXiv* **2024**. [CrossRef]
29. Blake, S.; Siddharthan, A.; Nguyen, H.; Sharma, N.; Robinson, A.M.; Elaine, O.; Darvill, B.; Mellish, C.; Van Der Wal, R. Natural language generation for nature conservation: Automating feedback to help volunteers identify bumblebee species. In Proceedings of the COLING 2012, Mumbai, India, 8–15 December 2012; pp. 311–324.
30. Cascaes Cardoso, M. The Onboarding Effect: Leveraging User Engagement and Retention in Crowdsourcing Platforms. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 263–267. [CrossRef]
31. Madeira, R.N.; Germano, H.; Macedo, P.; Correia, N. Personalising the User Experience of a Mobile Health Application towards Patient Engagement. In *Procedia Computer Science, Proceedings of the 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018)/The 8th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2018)/Affiliated Workshops, Leuven, Belgium, 5–8 November 2018*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 141, pp. 428–433. [CrossRef]
32. Gundelund, C.; Arlinghaus, R.; Baktoft, H.; Hyder, K.; Venturelli, P.; Skov, C. Insights into the users of a citizen science platform for collecting recreational fisheries data. *Fish. Res.* **2020**, *229*, 105597. [CrossRef]
33. Dechert, M. Implementation and Evaluation of a Chatbot to Crowdsource Geotagged Images to Detect Mosquito Breeding Sites. Master's Thesis, Universität Bremen, Bremen, Germany, 2019.
34. Tavanapour, N.; Poser, M.; Bittner, E. Supporting the Idea Generation Process in Citizen Participation—Toward an interactive system with a Conversational Agent as Facilitator. In Proceedings of the Twenty-Seventh European Conference on Information Systems (ECIS2019), Stockholm and Uppsala, Sweden, 8–14 June 2019.
35. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.t.; Rocktäschel, T.; et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, Vancouver, BC, Canada, 6–12 December 2020; ACM: Red Hook, NY, USA, 2020.

36. Yu, K.; Berkovsky, S.; Taib, R.; Conway, D.; Zhou, J.; Chen, F. User Trust Dynamics: An Investigation Driven by Differences in System Performance. In Proceedings of the 22nd International Conference on Intelligent User Interfaces, IUI '17, Limassol, Cyprus, 13–16 March 2017; ACM: New York, NY, USA, 2017; pp. 307–317. [CrossRef]
37. Samhale, K. The impact of trust in the internet of things for health on user engagement. *Digit. Bus.* **2022**, *2*, 100021. [CrossRef]
38. Khorashadizadeh, H.; Amara, F.Z.; Ezzabady, M.; Ieng, F.; Tiwari, S.; Mihindukulasooriya, N.; Groppe, J.; Sahri, S.; Benamara, F.; Groppe, S. Research Trends for the Interplay between Large Language Models and Knowledge Graphs. In Proceedings of the VLDB 2024 Workshop: The International Workshop on Data Management Opportunities in Unifying Large Language Models + Knowledge Graphs (LLM+KG), Guangzhou, China, 26 August 2024.
39. Ezzabady, M.; Ieng, F.; Khorashadizadeh, H.; Benamara, F.; Groppe, S.; Sahri, S. Towards Generating High-Quality Knowledge Graphs by Leveraging Large Language Models. In Proceedings of the 29th Annual International Conference on Natural Language & Information Systems (NLDB 2024), Turin, Italy, 25–27 June 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.