

Università  
della  
Svizzera  
italiana

Faculty  
of  
Informatics

Bachelor Thesis

June 11, 2025

# How do children search?

A tool to support researchers in understanding how children search for information online

Costanza Rodriguez Gavazzi

---

## *Abstract*

As children increasingly engage in web searching activities both at home and in educational settings, understanding how they search for information has become a key topic in both Information Retrieval (IR) and Human-Computer Interaction (HCI). However, most search interfaces are designed for adults and fail to address the cognitive and emotional needs of younger users. This thesis presents an interactive, web-based game designed to investigate how children search for information using either a traditional Search Engine (SE) or a Large Language Model (LLM). The game aims to be both engaging for children and practical for researchers. It logs structured interaction data, such as queries, tool preferences, session length, and user responses, and offers easy export of session logs for later analysis.

The project builds upon a previous prototype, making the system more modular, reusable, and suitable for real-world studies. Initial feedback from education professionals suggests the tool is promising as a research instrument, but still has room for improvement with respect to the User Interface (UI). This work provides a foundation for future iterations of the tool and supports the research community's need for structured, quantitative data on children's online information-seeking behavior.

**Keywords:** Information Retrieval; Human-Computer Interaction; User Interface; Search Engine; LLM; Children;

---

## **Advisor:**

Monica Landoni

## **Co-advisor:**

Diletta Micol Tobia

---

Approved by the advisor on Date:

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Background and Related Work</b>	<b>2</b>
2.1	Foundation . . . . .	2
2.2	Motivation . . . . .	3
<b>3</b>	<b>Requirements</b>	<b>4</b>
3.1	Data for Research Use . . . . .	4
3.2	UI Insights from Literature . . . . .	5
3.3	Functional Requirements . . . . .	5
<b>4</b>	<b>System Design</b>	<b>8</b>
4.1	Overview . . . . .	8
4.2	Structure . . . . .	8
4.3	APIs . . . . .	9
4.4	Technologies Used . . . . .	9
4.5	Research Data Collection . . . . .	9
<b>5</b>	<b>Usage Scenarios</b>	<b>10</b>
5.1	Experts' Feedback . . . . .	13
<b>6</b>	<b>Conclusions</b>	<b>15</b>
6.1	Future Work . . . . .	15
<b>A</b>	<b>Exported Data Example</b>	<b>17</b>

# 1 Introduction

In recent years, children's interactions with search technologies have attracted growing attention in both IR and HCI. As the Internet becomes more common in schools and homes, children are learning to search for information on their own. However, their needs and behaviors are different from those of adults. Commercial SEs like Google and Bing are not built with children in mind, and they often overlook key aspects of children's cognitive and emotional development.

At the same time, LLMs such as Chat-GPT<sup>1</sup> and Gemini<sup>2</sup> have become increasingly popular among the general population. Although these models were not originally designed to replace SEs, many people now use them to look up information they would otherwise search for online [15, 7, ?]. The single response offered by an LLM is sometimes seen as more convenient than browsing a list of search results [18, 4, 22]. Because of this, LLMs and SEs are often perceived as interchangeable, even though they work very differently [4, 14, 22, 23]. Such tools are also quickly evolving: new models include web browsing capabilities, allowing them to access real-time information. These systems combine LLMs with live web access<sup>3</sup> to provide up-to-date answers, further closing the gap between conversational assistants and traditional SEs [23, 17, 21, ?].

Previous research has shown that children struggle to evaluate search results effectively, particularly when results are presented in dense or complex formats. Moreover, their perception of what is relevant often differs from adult expectations and may be influenced by emotion, position in the list, or visual elements [11, 12].

This project builds on Savoia's 2025 thesis [16], which proposed a game to observe children's search behavior using both a traditional SE and a chatbot-like LLM interface. Because the original prototype was a proof of concept, it did not implement a data collection feature. This thesis rebuilds that tool with the goal of creating a functioning, reusable, and extensible prototype that can be used in real research scenarios and contribute to the development of child-friendly search tools.

To do this, the tool was designed as an interactive web-based game in which children answer questions framed in a neutral, positive, or negative emotional tone by choosing between two search tools. Their choices, search behavior, and answers are logged to generate structured data for research purposes. This allows to later analyze aspects such as tool preference, emotional influences, and performance across tasks.

This report is structured as follows:

- Section 2 reviews background research and related work, providing motivation for the project.
- Section 3 explains how the requirements were defined and lists them.
- Section 4 describes the technical architecture of the system.
- Section 5 shows typical usage scenarios.
- Section 6 concludes the report with a discussion of the directions for future work.

## 2 Background and Related Work

The study of how children search for information online is a growing area that connects education, IR, and HCI. The goal of this project was to develop a research tool that could collect interaction data from children using both traditional SEs and their Search Engine Results Page (SERP) and LLMs. In particular, the desire to create something that could be used, improved, and extended in the future was pivotal in guiding the work presented here. Unlike other projects focused mainly on improving User Experience (UX) or UI, this project was designed as a flexible platform to support various research questions on children's search behaviors. At the same time, the tool had to be fully functional and ready to use in real research scenarios, so an engaging and appropriate UI for younger users was still a key requirement.

### 2.1 Foundation

The starting point for this project was the bachelor's thesis by Savoia [16]. In her work, she proposed an interactive game designed to observe how children search for information. The interface was structured around a group of

---

<sup>1</sup><https://openai.com/chatgpt>

<sup>2</sup><https://gemini.google.com>

<sup>3</sup><https://openai.com/index/introducing-chatgpt-search/>

six islands, where each island represented a question framed with a specific emotional tone (positive, neutral, or negative). The children were allowed to choose how they wanted to answer the question, using a familiar web search interface or a chatbot-style LLM. Her work demonstrated that this narrative approach could encourage natural interactions and meaningful behavior from children.

Although her prototype successfully demonstrated the idea, it was not built with the core feature of data extrapolation, nor with reuse or extensibility in mind, but rather as a proof of concept. This project rebuilt the system from scratch to make it modular, maintainable, and suitable for future research experiments. Still, it kept the core idea of game play, the use of emotionally charged questions, and flexibility in tool choice to support various search strategies.

## 2.2 Motivation

A review of the literature highlighted the need for a research tool tailored to children’s needs. Children are frequent users of SE, especially in school settings, but often struggle to find relevant information efficiently [1]. Most commercial tools like Google or Bing are designed for adults [1, 10], and there is no single search interface that fits all users, especially young people with different cognitive and emotional needs [10].

In the IR research community, there is a clear call for child-friendly systems that are more suitable in educational contexts [13]. However, building such systems is challenging. Designing for children means taking into account their specific abilities: cognitive, technical, and emotional [5]. In many cases, children say they want one thing, but their actual behavior shows something else. This is why relying only on interviews or post-task questionnaires can be misleading [1].

Besides, children are often left out of mainstream IR research and there is a lack of reliable data on how they really use search tools [16]. Researchers have highlighted the need for dedicated datasets, experimental tools, and evaluation methods designed for children [11].

Observing how children search and interact with systems, we can gather important insights that help improve the design of future tools [16]. These insights can also guide the development of features that help children, for example, by offering visual cues to highlight relevance or reduce confusion [10]. Data from child search behavior can also support many different research directions, including how emotions impact search performance, how children respond to positive or negative task formulation, or how task complexity affects engagement [13, 11]. In particular, it is also interesting to consider how negative emotions are handled. Some research suggests that tracking negative user emotions could support better filtering, while others warn that hiding negative content could limit learning opportunities [13].

A game-based tool, like the one built in this project, can collect structured data such as query logs, number of queries, time spent, and user clicks [1]. It can also capture richer information on how children switch between tools or how many queries they need before answering. This kind of data would be difficult to gather with traditional observational methods or surveys, as children are reported to be unreliable sources when it comes to describing their own search behavior and preferences [1].

The strength of this approach comes from the benefits of game-based learning, which helps keep children engaged and motivated [2]. When children play a game, they tend to act more naturally and stay focused, making it easier to collect useful data on their behavior [2]. Compared to traditional learning environments, which can feel serious or stressful, games create a more relaxed space where children feel free to explore [2].

By designing the tool as a digital game, we also benefit from interactive elements, storytelling, and visuals that make the experience fun.

### 3 Requirements

#### INTRO GIOCO

This project did not begin with a fixed list of specifications. Instead, requirements emerged iteratively through exploration of the research domain, analysis of related literature, and discussions with advisors, both researchers in IR and, therefore, stakeholders in this design space. The work builds on Savoia's thesis [16], which included an initial prototype of the game that was tested in a pilot session with children. From that experience, the feedback from the advisors and what we found in the literature, a new version of the game came to life, one that is more functional, scalable, and better suited to support different research questions in IR and HCI, while also being engaging and usable for children.

The main objective was to implement a fully functional and reusable prototype that could collect detailed data on how children use traditional SE and LLM, while also being engaging and usable for children. The project extends Savoia's prototype [16], which demonstrated the potential of a game-based framework to study children's search behavior, but did not actually integrate the aspect of user data collection. To structure the design, the dual purpose of the tool was fundamental: relevance for researchers aiming to study search behavior, and a pleasurable and engaging experience for children. From this core duality, the following questions guided the requirements definition process: To structure the design, the following questions guided the requirement definition process.

- What kind of data do researchers want to collect?
- Can a game-based interface be used to collect these data effectively?
- What kind of game design is both engaging and suitable for structured data collection?
- How can the tool be both usable for researchers and appealing to children?
- How can the system be modular, extensible, and maintainable for future research?
- How can it adapt to different experimental setups or research contexts?

These reflections acted as a guide to explore the literature, which then led to a concrete list of functional and non-functional requirements, discussed below.

#### 3.1 Data for Research Use

Research has shown that a variety of data is needed to understand children's search behaviors. This includes both quantitative metrics and qualitative observations. Although qualitative observations can offer several advantages in designing for children, this project's main focus was on quantitative metrics to enable the identification of patterns or relationships among behaviors. Analysis of these patterns can help researchers identify the potential roles of searchers among children [9]. Understanding these distinct roles that children play when searching is a key element in characterizing their behavior. Previous research [8] defined seven roles for children searching in a home setting, based on qualitative data such as interviews and observations. More recent research [11] aimed to understand the search roles of children, specifically in the classroom, and to see if they could be inferred using a quantitative approach based on performance indicators from search logs and teacher evaluations. By analyzing quantitative data such as click accuracy, session length, and grades, researchers were able to identify patterns that corresponded to several of the original roles [11]. These findings highlight that both qualitative and quantitative data are fundamental to understanding children's search behavior.

Inspired by the approach of [11], we extended the analysis by identifying the following types of data as relevant.

- **Query Logs:** Query text, number of queries, and query term counts [1, 16].
- **Performance Indicators:** Session length, click counts, and rank depth [11].
- **Task Outcome Data:** Final user responses, with the idea that scoring based on teacher rubrics or predefined relevance criteria could be done later when analyzing the gathered data [11].
- **Tool Preference:** Children's selection patterns between SE and LLM [16].
- **Timing Data:** Time spent on each question, delays before interaction, and time spent on web pages [16].

### 3.2 UI Insights from Literature

Designing a UI involves understanding how users interact with a digital system through visual and interactive elements. In this project, the UI was not the sole focus but because the target users were children, it required some careful thought. Designing UI for children presents challenges: it must be intuitive, engaging, and customized to their cognitive and emotional development. At the same time, the interface should allow adult stakeholders, such as teachers and researchers, to take any necessary actions, such as setting up the game and downloading the gathered data.

Considering the dual purpose of the tool, the following insights from the literature informed the UI design:

- There is no one-size-fits-all SERP interface suitable for children [10].
- Children prefer playful and engaging tools, and emotional design improves product recognition and user satisfaction [5, 3].
- Humanized design elements such as avatars, animations, and storytelling are particularly effective for younger users [3, 5].
- Participatory design methods show that children prefer familiar, usable UI elements but appreciate fun, cartoonish visuals [20, 12].
- Children prefer to interact with a tool that they perceive to be similar to the real thing, rather than a simplified version [1].
- Based on their age and cognitive abilities, children require different levels of scaffolding<sup>4</sup> and support in search tasks [5, 20, 3]. In particular:
  - **Ages 6 to 8:** Interfaces should use simple, familiar words, and reward systems are effective in maintaining engagement. Bright colors are still encouraged [3].
  - **Ages 9 to 12:** These users value autonomy and prefer interfaces that offer control rather than instructions. Feedback should be informative rather than directive. Colors can be more subdued, such as green, gray, and navy tones. They are generally more skilled at navigating websites and handling smaller UI elements [3].

### 3.3 Functional Requirements

Based on the insights from the literature and discussions with the advisors, the following functional requirements were defined for the system:

#### Language and Initialization

- The language of the game should be modular and easily changeable to be able to test the game with children from different countries.
- A landing page must appear before the game starts to allow teachers or researchers to explain the rules.
- The game must be reset when a new language is selected.

#### Session Management

- Each session must be identified by a unique user code provided by the teacher.
- The game must start only after the user inputs a valid code.
- The game must persist its state on page refresh using local storage.
- Researchers must be able to download session data at any time, even if the game is incomplete.
- When downloading data, the system must request a password to protect the export and prevent children from tampering with it.

---

<sup>4</sup>Scaffolding refers to a form of support provided in design, particularly for children [5, 11]. In app design, providing scaffolding involves conveying the objectives of the app to children in an understandable way (e.g., interactive prompts, visual cues) [5, 11]

## Game-play Logic

- The game must include six clickable islands, presented in a randomized order per session.
- Islands can be clicked in any order.
- Each island represents a straightforward question with a specific emotional framing: positive, neutral, or negative.
- Once answered, an island becomes inactive and cannot be clicked again.
- Inactive islands must show a different cursor, remove hover effects, and a message should appear when trying to click on them.
- Each completed island increases the score by 10 points (maximum 60).
- The current score must be shown to motivate the user.

## Search Tool Interfaces

- Clicking on an island must open a question page where the user chooses between two different search tools: Google or Gemini.
- Users must be allowed to switch between the two tools freely before submitting an answer.
- The Google interface must visually mimic a real SE and display the top 10 results.
- The Gemini interface must mimic a realistic chatbot interface.
- Clicking on a Google result must open the link in a new browser tab.

## User Interaction Tracking

The system must record **per-island data**:

- Question
- Sentiment of the question
- Time of the first click on the island
- Time of the answer submission for the island
- A list of objects containing information about each query or prompt performed by the user, in particular:
  - If the query was made using Google or Gemini
  - The text of the query or prompt
  - The number of query terms
  - The answer provided by the tool of choice - either a list of SERP results for Google or the response from Gemini.
  - When saving SERP results, the system must record:
    - \* The title of the result
    - \* The snippet of the result
    - \* The position of the result in the SERP
    - \* The order in which the result was clicked
    - \* The time spent on the page corresponding to the result
- The user's final submitted answer

The system must also record **per-session data**:

- Language of the game
- A code to identify the session, provided by the teacher

- Start time and finish time of the session
- Session length
- The order in which the islands were completed
- The order in which the islands were clicked
- Total number of clicks in the session
- The time before the first click in the session
- The final score of the session

#### **UC and Engagement**

- The game must allow revisiting of islands before submission.
- The emotional tone must be visually represented with colors or icons.
- The UI must balance playful and realistic design to reflect real-world behavior of children when interacting with games.
- Upon completion, a thank you page must show the user's final score.
- The background includes animations to enhance engagement.

## 4 System Design

### 4.1 Overview

The game is implemented as a single-page web application developed using React, a JavaScript framework. The architecture follows a standard component-based design, where each part of the UI is a block, potentially reusable. All logic, including game progression, timing, data collection, and API communication, is embedded directly in the relevant components or custom hooks.

There is no back-end: the entire game runs in the user's browser. State is managed using React's Context API and saved to `localStorage` to support persistence across sessions. When needed, the app communicates with external APIs to retrieve results from a SE (Google) or a LLM (Gemini).

### 4.2 Structure

The main components of the application represent the main stages of the game and are shown in Fig. 1. The flow begins at `ChooseLanguagePage` and proceeds through `GameStartPage`, `MapPage`, `ChoicePage`, and `QuestionPage`, concluding at `FinishPage`.

Although some transitions are one-way (for example, from start to map), the design supports two-way navigation between `MapPage`, `ChoicePage` and `QuestionPage`, allowing users to explore and revisit islands freely. The state and data collected during navigation are tracked using a centralized context. The `QuestionPage` triggers asynchronous calls to either the Google Search API or the GoogleGenerativeAI API, based on the user's choice.

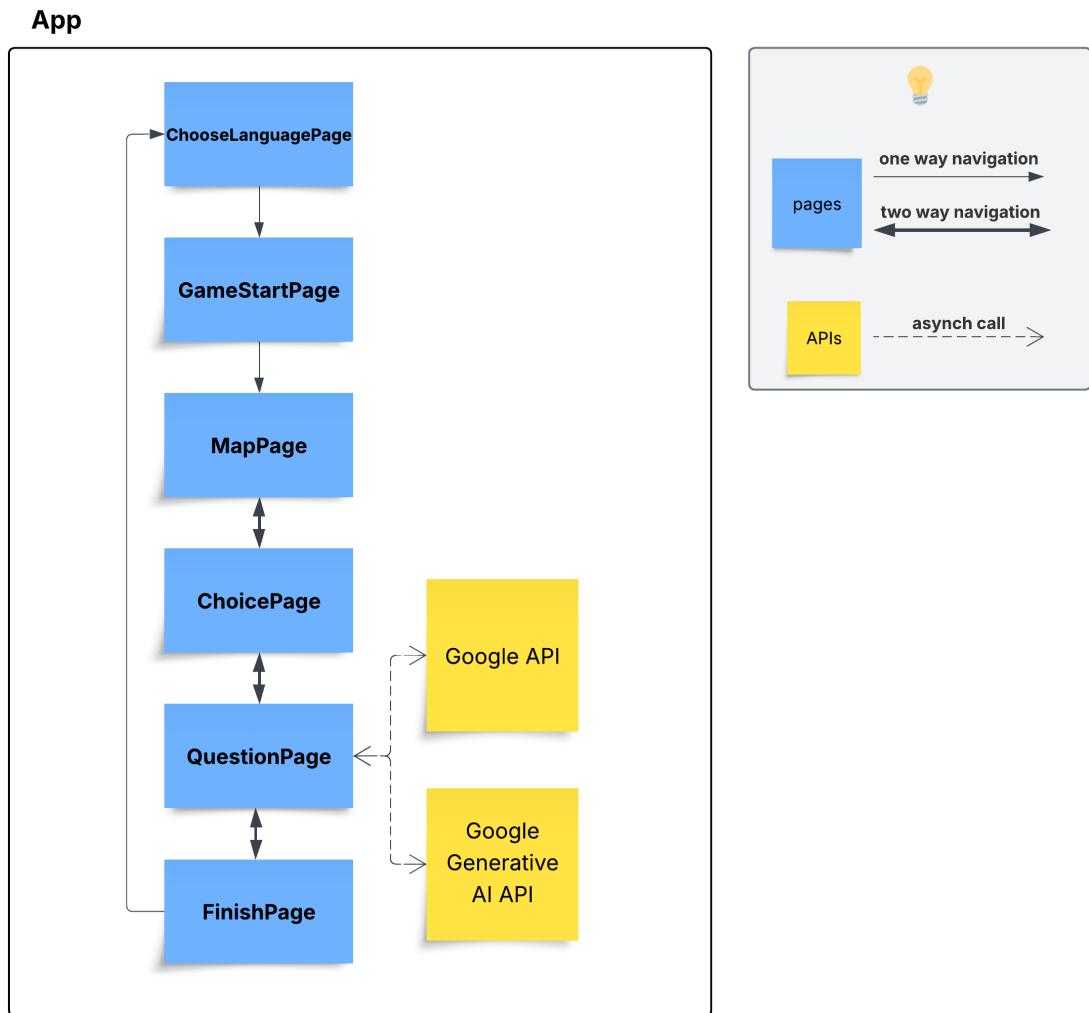


Figure 1. Structure of the application.

### 4.3 APIs

The game uses two external services in order to retrieve the information from the user's prompts, selected based on the player's choice:

- **Google Search API** Performs a custom Google search and retrieves the top 10 results for a given query. When users select the *Google* option in ChoicePage, this API returns SE results, each including metadata such as title, snippet, rank, whether it was clicked, click order, and time spent on the result.
- **GoogleGenerativeAI API (Gemini)** Sends the user's prompt to the Gemini LLM and returns a generated answer. This is selected when the player chooses the *Gemini* option. Responses are stored in the local state alongside metadata such as prompt length and timestamp.

### 4.4 Technologies Used

The application is built using the *Create React App* boilerplate, which provides a preconfigured setup for React applications. The main technologies used are as follows:

- **React**: Core framework for building the single-page application with reusable components and hooks.
- **JavaScript**: Used for writing all the game logic and component logic.
- **Mantine UI**: A React component library.
- **Plain CSS**: Used for styling components, organized in modular files like `map.css`, `question.css`, and more.
- **Google Custom Search API**: Used to perform keyword-based queries when users choose the source 'Google'.
- **GoogleGenerativeAI API (Gemini)**: Provides LLM-based responses to user prompts when the "Gemini" option is selected.
- **React Context + Hooks**: For managing shared state (e.g., game data, API responses).
- **Browser LocalStorage**: Save session data, including game state and interaction logs.
- **Environment Variables (.env)**: API keys are stored in a private `.env` file and are accessed at runtime.

Custom utility hooks such as `useWindowDimensions` and `useAutoScale` dynamically adapt the UI layout. The global game state is managed with React's Context API in the `GameStateContext` component.

### 4.5 Research Data Collection

The game collects structured data to support research on how users interact with different information sources (SE vs. LLM). All data is stored locally and can be exported at the end of the session.

- **Session Metadata:**
  - `userCode`, `startTime`, `finishTime`, `sessionLength_seconds`
  - `totalClicksInSession`, `islandClickOrder`, `islandCompletionOrder`
- **Per-Island Data:**
  - `question`, `sentiment`, `userAnswer`
  - `choiceForAnswer`: Logs whether the user chose Google (0) or Gemini (1)
  - `Queries`: Contains detailed records of each query:
    - \* `query`: Text of the input, `AI`: Flag (0 for Google, 1 for Gemini), `numberOfQueryTerms`, `answer`: Either a plain string (for Gemini) or a list of search results (for Google), each entry made of title, snippet, position, click status, click order, time spent on result

All of this information can be exported through the `exportData()` function and analyzed to study user behavior, preference trends, and differences between IR strategies. An example of exported data is shown in Appendix A.

## 5 Usage Scenarios

This section illustrates how researchers or teachers (stakeholders) can set up the game and download the data, and how a child interacts with it during a typical session.

### Step 1: Language Selection

The game begins with the language choice screen (Fig. 2). The stakeholders select their preferred language before continuing, setting up the game for the child. A hover effect provides visual feedback (Fig. 3).

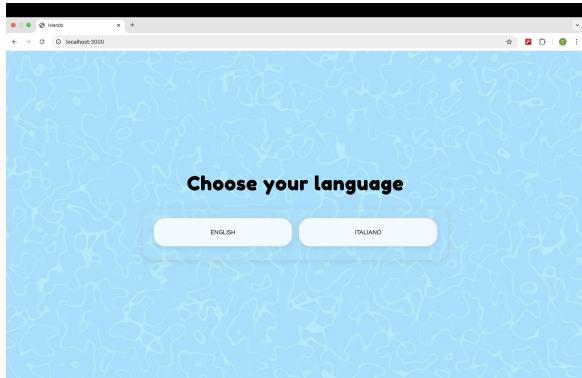


Figure 2. Language choice screen.

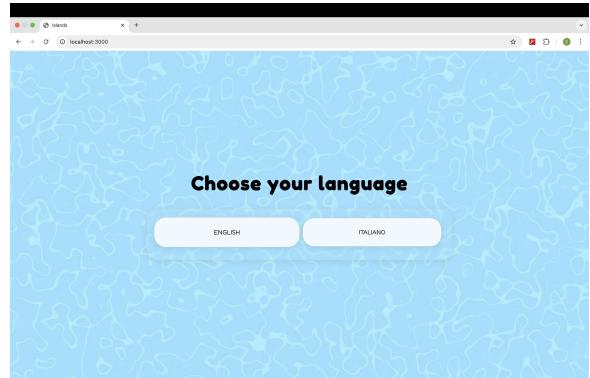


Figure 3. Hover effect on language choice.

### Step 2: Session Start

After an adult has chosen the language, the child sees the start screen of the game (Fig. 4). A hover animation was implemented, guiding them to start (Fig. 5).

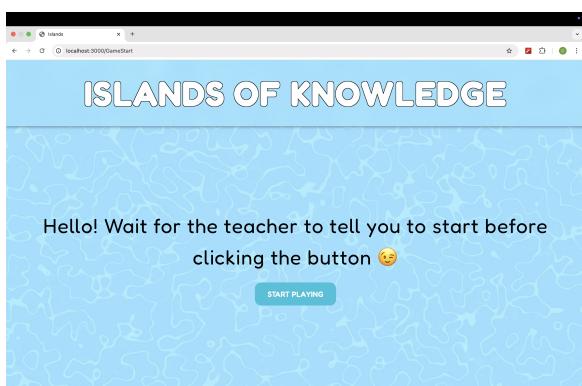


Figure 4. Game start screen.

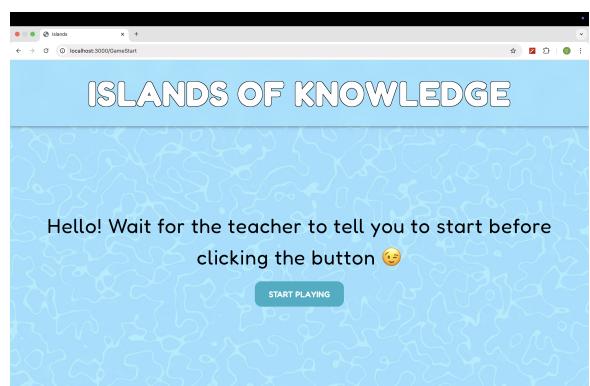
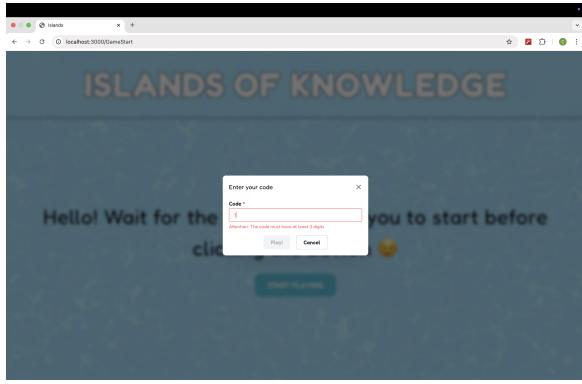


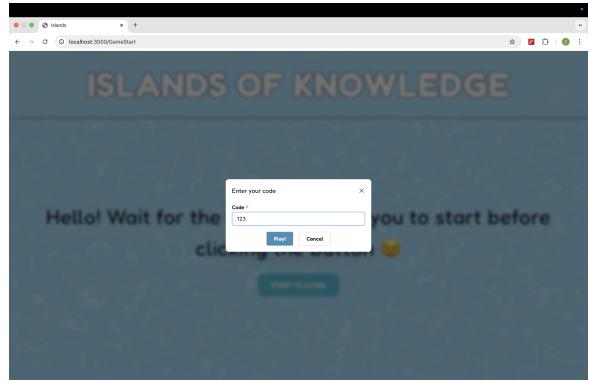
Figure 5. Hover effect on start button.

### Step 3: Entering a Code

The system requires the child to enter the session code they received prior to the start of the session (Fig. 6), which is validated (Fig. 7). This ensures that all recorded data is associated with a specific session.



**Figure 6.** Session code entry screen.



**Figure 7.** Successful code validation.

#### Step 4: The Island Map

The core component of the game corresponds to the map page (Fig. 8). Each of the six clickable islands represents a unique search task. The hover feedback (Fig. 9) helps guide the attention of the children.



**Figure 8.** Island map with clickable islands.



**Figure 9.** Hover effect on islands.

#### Step 5: Tool Choice

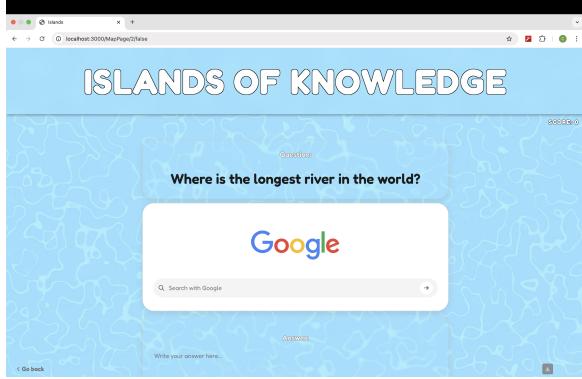
After selecting an island, the child reaches a tool selection page (Fig. 10) where they choose between a traditional SE (Google) and a LLM (Gemini).



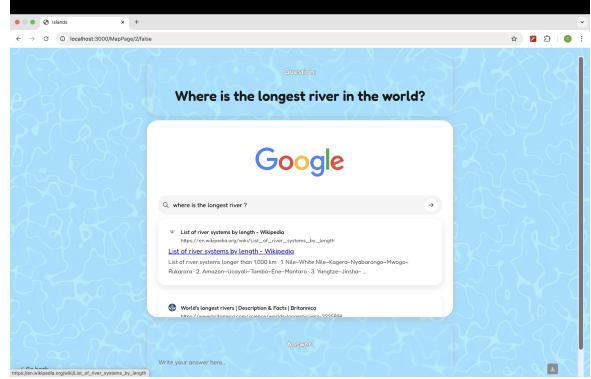
**Figure 10.** Tool choice screen.

#### Step 6: Using Google

If the child selects Google, they are presented with a simulated SE landing page (Fig. 11). The interface and behavior mimic real-life scenarios (Fig. 12). Clicking on links opens them in a new tab.



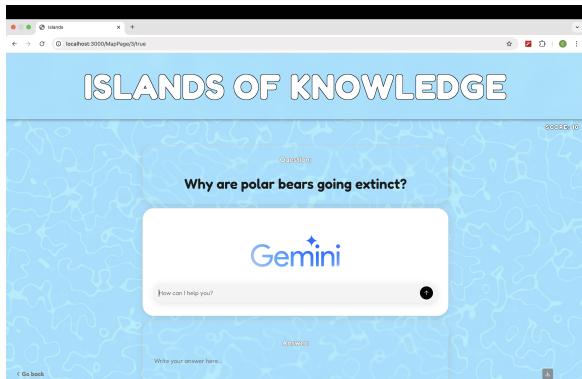
**Figure 11.** Simulated Google landing.



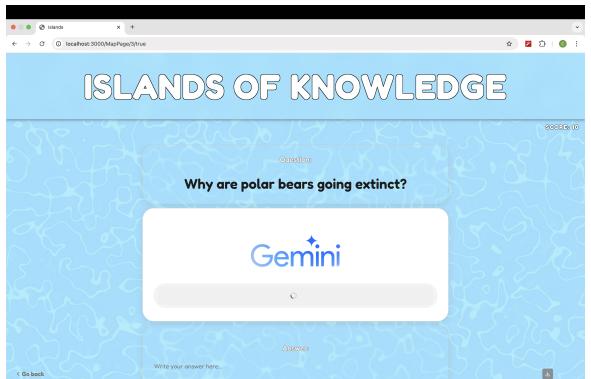
**Figure 12.** Google search results with clickable links.

## Step 7: Using Gemini

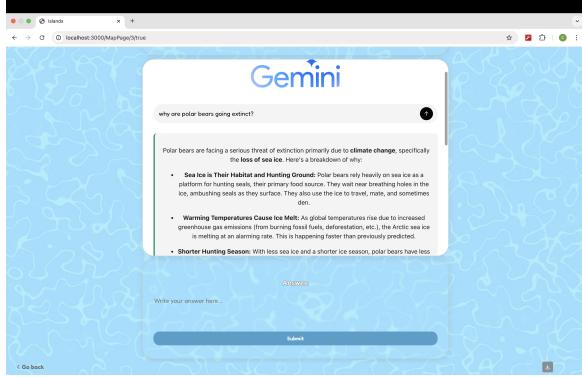
Alternatively, children can choose Gemini, where a chat-like interface appears (Fig. 13). A waiting screen mimics the LLM response delay after sending a prompt (Fig. 14), followed by the result output (Fig. 15).



**Figure 13.** Gemini chat interface.



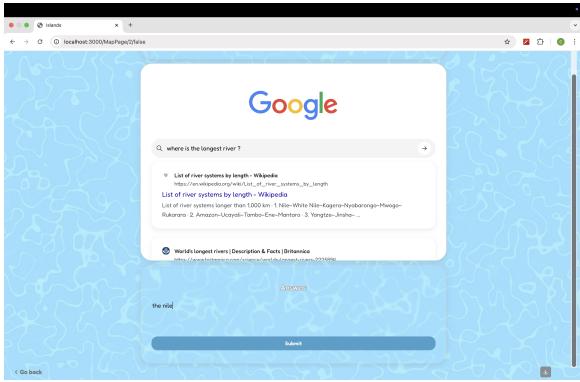
**Figure 14.** Waiting screen for Gemini response.



**Figure 15.** Gemini response output.

## Step 8: Submitting an Answer

After exploring the results, the child can type the final answer (Fig. 16). Once submitted, the user is redirected to the map page, and the island becomes inactive. If the child attempts to revisit a completed island, a warning is shown (Fig. 17).



**Figure 16.** Final answer submission screen.

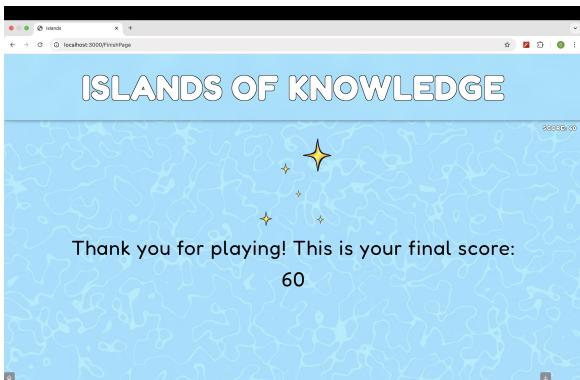


**Figure 17.** Can't click on a completed island.

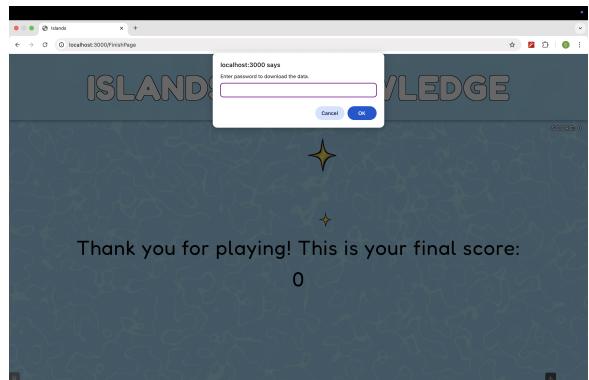
### Step 9: Final Page and Data Export

After completion of all six islands, the child sees their final score (Fig. 18). Each completed island contributes 10 points, with a maximum of 60.

To end the session, a password-protected download prompt is shown (Fig. 19), allowing the teacher or researcher to save all session data locally. It was a design choice to use the browser's prompt functionality instead of a custom modal, to discourage children from trying to tamper with the data export.



**Figure 18.** Final score screen after completing all islands.



**Figure 19.** Password-protected data download prompt.

### 5.1 Experts' Feedback

Three primary school teachers with experience in education and technology reviewed the tool. They were shown a video recording of simulated gameplay with a voiceover explanation, which served as a demo of how the game works. These teachers regularly conduct computer-related lectures in classroom settings in Ticino. Here, children typically begin using Google in third grade under close guidance. At this stage, the experts reported on the children's struggle with navigating the SERP, switching between tabs, and identifying which results to click on. They provided valuable feedback on the design and usability of the tool, which should be discussed during future iterations of development. Please note that the feedback should be considered as input to be discussed and further analyzed, rather than as definitive requirements to be implemented.

1. The overall visual design was considered slightly outdated and overly childlike, more appropriate for younger students (third or fourth grade). Two experts suggested a more modern and polished look to better suit the target age group.
2. Regarding the first screen, the introductory message was described as too long. Since children tend to click instinctively, they might skip important instructions. To address this, two alternative solutions were proposed:
  - (a) Split the prompt into two screens, one with a warning to wait for the teacher's instruction, and a second screen with a clickable button.
  - (b) Using a larger font and shortening the text (e.g., "WAIT! When you hear 'ok', click").

3. Although appreciated in terms of idea and icons used<sup>5</sup>, the screen featuring clickable islands was deemed not intuitive. The experts noted that it did not encourage interaction and that some children might not realize that the elements were clickable. Additionally, the use of happy and sad face icons was discouraged, as they could lead children to assign emotional meaning or personify the interface unnecessarily.
4. For the question screen, both font size and type were mentioned. They believed that the text should be larger and use a highly readable font such as Arial or certified dyslexia fonts. Furthermore, the question area was suggested to be visually highlighted (e.g., with a colored background) to make it stand out from the rest of the interface.
5. The use of commercial logos was also criticized. This was flagged both for ethical reasons and for inconsistency: although a Google-like logo was shown, the following screen was considered not an accurate replica of the Google or Gemini interface. As a workaround, it was recommended to label the search interface simply as “Search Engine” while mimicking the Google look through color and font style.  
*Note:* The primary goal of the tool is to capture realistic data on children’s search behavior. To support this objective, the design intentionally mirrors real-world tools as closely as possible. Furthermore, since the game integrates the official Google and Gemini APIs, the search results displayed to the children are retrieved directly from these services. For this reason, the decision was made to include the actual logos in order to reflect the real context in which children typically conduct searches.
6. The score system also received mixed feedback. It was suggested to give teachers the option to enable or disable the display of points. For children with lower self-esteem, visible scores could be discouraging. In contrast, group dynamics could benefit from visible scoring as a form of motivation. A teacher explicitly reported that scores fascinate and stimulate children, who appreciated them. If scores are shown, they should be designed with engaging, attention-grabbing visuals.
7. The simulated Google interface needed refinement. The space for displaying the question was perceived as difficult to notice, and the answer input field was not clearly interactive. Improving layout cues would help clarify the expected user action.
8. The final screen’s thank-you message was considered too long. A concise alternative such as “Great job! You earned X points!” was recommended, particularly if the scoring feature remains active.
9. One teacher raised concern about the lack of scaffolding during the research phase of the game, as younger students may require structured support to conduct meaningful searches. It was suggested to provide curated websites alongside the guiding questions, ensuring that students are able to find suitable information independently without being overwhelmed.  
*Note:* As previously stated, the objective of the tool is to collect realistic data. Adding scaffolding during the retrieval process would reduce the ecological validity<sup>6</sup> of the results, as it would no longer reflect how children interact with actual tools in real world settings.
10. The choice page was appreciated by the experts and considered a good idea to encourage decision-making during the activity.

These insights are essential to improve both the usability and accessibility of the tool in future versions.

---

<sup>5</sup>One expert recalled a similar island-based activity on the platform [www.kodable.com](http://www.kodable.com)

<sup>6</sup>Ecological validity is the degree to which research findings or assessments reflect, resemble, or predict real-world behaviors and contexts [6, 19].

## 6 Conclusions

This project focused on the design and development of a tool to support researchers collecting user data to study how children interact with different IR systems. The tool, designed in collaboration with researchers and teachers and built as an interactive web-based game, allows children to choose between a traditional SE and a chatbot-style LLM to answer emotionally framed questions. All interactions are logged in a structured way, allowing the collection of data to support research.

The work originated from the limitations identified in Savoia's thesis [16], particularly the need for data collection features. Although her work laid the foundation for a possible approach to studying search behavior, it did not provide a usable, extensible, or customizable tool for research purposes. This project rebuilt the system with the goals mentioned above in mind.

Several contributions were achieved:

- A fully functional prototype was implemented.
- A data model was developed to log user interactions in detail, capturing a variety of metrics relevant to IR and HCI research.
- The UI was designed to balance child-friendly visuals with realistic representations of search interfaces, aligning with guidelines from the literature.
- Feedback from experts in education and child-computer interaction was gathered to inform future improvements and highlight potential usability issues.

The project provides a functional addition to the current state-of-the-art in the IR and HCI research domains, particularly in the context of studying children's search behavior. The tool is designed to be modular and extensible, allowing future adaptations to different research contexts or experimental setups.

### 6.1 Future Work

Although the current prototype is functional and is already suitable for small-scale research studies, further improvements can be made in future iterations to improve usability, data collection, and scalability.

#### UI and Usability Testing

The current interface was designed based on the insights of the literature and internal feedback, but because the design interface could have been a whole project on its own, it can be improved. Furthermore, it has not yet been tested with children in real-life settings. One key area for future work is to conduct formal usability testing sessions with children of different age groups. These sessions would help identify which parts of the UI are intuitive, which are appealing, and which are confusing or too complex. For example, visual feedback, such as hover animations, tool choices, or iconography, may need to be adjusted based on children's preferences and cognitive abilities. Additionally, working with experts in child-computer interaction and UX/UI could help refine the design to better suit the target audience.

#### Extending Data Collection

The current version of the system already collects a wide set of interaction data, including query logs, tool switches, time spent, and submitted responses. However, there is room for improvement and expansion. Future versions could include additional behavioral signals, such as

- Mouse movement and hover times to infer hesitation or curiosity.
- Scroll depth within the SERP.
- Optional feedback prompts for children to self-report satisfaction or perceived difficulty.

#### Improving Efficiency and Scalability

In the current version of the system, all interaction data is stored locally in the browser and must be manually exported at the end of each session. Although this solution is suitable for small-scale studies or pilot experiments, it introduces several limitations when conducting larger studies.

Managing data from multiple devices becomes time-consuming, and there is a risk of losing session data if a participant closes the browser before exporting. Relying on browser-based password prompts also provides only minimal protection and may not be ideal in more formal research contexts.

A possible solution could be to integrate a back-end infrastructure that stores the data in a centralized manner, allowing researchers to:

- Monitor sessions in real time, even across multiple classrooms or devices.
- Automatically collect and aggregate data without manual intervention.
- Minimize the risk of data loss due to user error or technical problems.

### **Architecture and Code Quality**

The current codebase is functional but there is always room for improvement. Future work could focus on separating the logic from the UI components, implementing a separate layer for game functionality, and using a more modular architecture. This refactoring would further improve maintainability and scalability.

### **Experts' Feedback**

The feedback from the experts highlighted several areas for improvement, particularly in the UI design and usability aspects. Future work should focus on discussing the feedback to enhance the overall UC. This includes refining the visual design to be more modern and polished, simplifying instructions, improving the intuitiveness of clickable elements, and ensuring that the interface is accessible to children with different cognitive abilities.

## A Exported Data Example

The following is an example of the data structure exported by the system at the end of a session. The data is structured in JSON format, which allows for easy parsing and analysis.

```
{  
    "sessionMetadata": {  
        "userCode": "12345",  
        "startTime": "2023-10-01T10:00:00Z",  
        "finishTime": "2023-10-01T10:15:00Z",  
        "sessionLength_seconds": 900,  
        "totalClicksInSession": 30,  
        "islandClickOrder": [0, 1, 2, 5, 3, 4, 1, 2, 5],  
        "islandCompletionOrder": [0, 3, 4, 1, 2, 5]  
},  
    "islandsData": [  
        {  
            "question": "What is the capital of France?",  
            "sentiment": "neutral",  

```

## References

- [1] M. Aliannejadi, M. Landoni, T. Huibers, E. Murgia, and M. S. Pera. Children's perspective on how emojis help them to recognise relevant results: Do actions speak louder than words? In *CHIIR '21: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, Canberra, Australia, March 14–19 2021.
- [2] M. S. Alotaibi. Game-based learning in early childhood education: a systematic review and meta-analysis. *Frontiers in Psychology*, 15:1307881, 2024. Received: 05 October 2023; Accepted: 20 March 2024; Published: 02 April 2024.
- [3] A. Amin, M. Shidujaman, and B. Wang. Improving hci on cognition for children with intelligent ui/ux. In *Conference on Human-Computer Interaction*, 2022.
- [4] K. M. Caramancion. Large language models vs. search engines: Evaluating user preferences across varied information retrieval scenarios. *ArXiv*, abs/2401.05761, 2024.
- [5] K. Chen. An interactive design framework for children's apps for enhancing emotional experience. *Interacting with Computers*, 2022. Advance Access published on 28 December 2022.
- [6] D. Dawson and T. Marcotte. Special issue on ecological validity and cognitive assessment. *Neuropsychological Rehabilitation*, 27:599 – 602, 2017.
- [7] R. R. Divekar, S. Guerra, L. Gonzalez, and N. Boos. Choosing between an llm versus search for learning: A highered student perspective. *ArXiv*, abs/2409.13051, 2024.
- [8] E. Foss, A. Druin, R. Brewer, P. Lo, L. Sanchez, E. Golub, and H. Hutchinson. Children's search roles at home: Implications for designers, researchers, educators, and parents. *Journal of the American Society for Information Science and Technology*, 63(3):558–573, 2012. First published: 18 November 2011.
- [9] M. Imazu, S. Nakayama, and H. Joho. Effect of explicit roles on collaborative search in travel planning task. In *Proceedings of the International Conference on Information Seeking and Retrieval*, 2017.
- [10] M. Landoni, M. Aliannejadi, T. Huibers, E. Murgia, and M. S. Pera. Have a clue! the effect of visual cues on children's search behavior in the classroom. In *Proceedings of the ACM Conference on Human Information Interaction and Retrieval (CHIIR)*, Canberra, Australia, March 14–19 2021.
- [11] M. Landoni, T. Huibers, M. Aliannejadi, E. Murgia, and M. S. Pera. Getting to know you: Search logs and expert grading to define children's search roles in the classroom. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR)*, 2021.
- [12] M. Landoni, T. Huibers, E. Murgia, M. Aliannejadi, and M. S. Pera. Somewhere over the rainbow: Exploring the sense for relevance in children. In *Proceedings of the ACM Conference on Human Information Interaction and Retrieval (CHIIR)*, 2021.
- [13] M. Landoni, M. S. Pera, E. Murgia, and T. Huibers. Inside out: Exploring the emotional side of search engines in the classroom. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*, Genoa, Italy, July 14–17 2020.
- [14] N. F. Lindemann. Sealed knowledges: A critical approach to the usage of llms as search engines. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 2023.
- [15] T. Mendel, N. Singh, D. M. Mann, B. Wiesenfeld, and O. Nov. Laypeople's use of and attitudes toward large language models and search engines for health queries: Survey study. *Journal of Medical Internet Research*, 27, 2024.
- [16] L. Savoia. A quest for knowledge - coding and developing an interactive game to trace children's online searching experience, Jan. 2025.
- [17] X. Shi, J. Liu, Y. Liu, Q. Cheng, and W. Lu. Know where to go: Make llm a relevant, responsible, and trustworthy searchers. *Decis. Support Syst.*, 188:114354, 2024.
- [18] M. Stadler, M. Bannert, and M. Sailer. Cognitive ease at a cost: Llms reduce mental effort but compromise depth in student scientific inquiry. *Comput. Hum. Behav.*, 160:108386, 2024.

- [19] Y. Suchy, L. A. DesRuisseaux, M. G. Mora, S. L. Brothers, and M. Niermeyer. Conceptualization of the term “ecological validity” in neuropsychological research on executive function assessment: a systematic review and call to action. *Journal of the International Neuropsychological Society*, 30:499 – 522, 2024.
- [20] M. Tapola, T. Mäkilä, N. Erdmann, and M. Mikkilä-Erdmann. Participating elementary school children in ui design process of learning environment: Case kidnet. In *Proceedings of the Conference on Human Factors in Computing Systems*, Turku, Finland, 2022.
- [21] P N. Venkit, P. Laban, Y. Zhou, Y. Mao, and C.-S. Wu. Search engines in an ai era: The false promise of factual and verifiable source-cited responses. *ArXiv*, abs/2410.22349, 2024.
- [22] T. Worledge, T. Hashimoto, and C. Guestrin. The extractive-abstractive spectrum: Uncovering verifiability trade-offs in llm generations. *ArXiv*, abs/2411.17375, 2024.
- [23] H. Xiong, J. Bian, Y. Li, X. Li, M. Du, S. Wang, D. Yin, and S. Helal. When search engine services meet large language models: Visions and challenges. *IEEE Transactions on Services Computing*, 17:4558–4577, 2024.