

1 Methods

For our study, we implemented a natural language interface in the form of a webshop that allows users to directly compare our approach of dynamically generated queries with a traditional static template. This setup lets users try both methods directly and compare how well they retrieve relevant search results. For our example webshop, we chose to focus on a single category of items to ensure a more controlled and meaningful analysis. We ultimately decided on a webshop offering dogs, as dogs are highly popular and provide a diverse range of characteristics for comparison. The interface was structured so that the displayed search results were divided into two distinct sections. On the left side, the results were generated using AI-driven, dynamically created search queries, while on the right side, the results were retrieved using a traditional, static query template, incorporating a semantic matching function, as observed on various websites. The user entered their search query through a shared search bar. We decided to apply the filtering functionality to both sections via a unified sidebar to avoid user confusion, even though it was not strictly necessary for the AI-driven side, as the constraints covered by the filters could also be expressed in natural language. This design enabled users to directly compare both approaches in real time. By reviewing both result sets side by side, participants could evaluate the effectiveness of each method and determine which approach best aligned with their personal search behavior and preferences.

1.1 Implementation of the Interface

Our initial approach for dynamically generating SQL queries was to use a local large language model. To further refine the performance, we intended to implement the DAIL schema. However, the main issue with using a local model was that it had to be loaded into the computer’s RAM, which limited us to smaller models due to hardware constraints. These smaller models, however, did not deliver the expected performance in terms of both the accuracy of the generated SQL queries and the time required to generate them. As a result, we explored the possibility of using an API-based solution. For this, we tested both the Claude API and the OpenAI API. After extensive testing, we found that the Claude API performed slightly better. Consequently, we decided to integrate the Claude API into our system. The API delivered such strong results for our interface that, due to time constraints and the additional complexity involved, we ultimately decided not to implement the DAIL schema. In the prompt we used to dynamically generate SQL queries, we provided the API with both the database schema and all possible attribute values. This allowed the AI to correctly interpret user input and generate the appropriate query, even when the input was vague or descriptive. Additionally, we included a "base query" in the prompt, which outlined a general structure for the SQL query that the AI would then refine and complete based on the user’s input.

1.2 User Study

In the course of our user study, a group of 13 participants was examined, consisting of 12 males and one female. The majority of the participants were currently enrolled in a Bachelor’s or Master’s program in Computer Science. Additionally, most had prior experience with artificial intelligence, either through academic studies or personal interest. Before the study began, all participants were informed about its primary objective, which was to evaluate the performance of the AI-driven search system in direct comparison to a conventional, template-based search mechanism. By conducting this study, we aimed to determine whether the AI system could produce superior results in product search. The participants were asked to answer a predefined set of questions (see Appendix) to assess their experience and the system’s performance. The user study was divided into three phases. In the first phase, participants were asked

a series of general questions before interacting with the interface. These questions aimed to gather insights into their past search behavior, including any difficulties or challenges they had encountered when searching for specific items and the consequences these issues had on their overall shopping experience. In the second phase, participants were introduced to our interface and given the opportunity to explore its functionality. To familiarize themselves with the system, participants were first asked to complete a set of predefined search tasks. These tasks aimed to provide an understanding of the interface's functionality and the variety of dogs and their attributes. They were allowed to conduct these searches freely, choosing between traditional static filters and natural language queries. The tasks were designed so that the first few could be solved relatively easily using filter settings, while the later ones were either very difficult or even impossible to accomplish with filters alone—for example, "OR" queries. This ensured that participants would engage with the AI-based system and recognize its unique and superior capabilities compared to the static template. Participants also had the option to refine or correct their queries if the initial results were unsatisfactory. After this guided introduction, they were encouraged to experiment with their own complex search queries using natural language. The final phase of the study involved a more detailed questionnaire designed to evaluate the AI-driven approach in comparison to the conventional model. To assess participant satisfaction with our interface and the AI-based approach, we included several questions in this part of the study, which were rated on a 5-point Likert scale. We also investigated participants' perspectives on the role of filtering mechanisms, exploring whether a model like ours could serve as a potential replacement for traditional filtering functionalities.