UNC Dorm Guide:
An Assistant for Freshman Housing Recommendations

Shreeya Sethi
University of North Carolina at Chapel Hill
shreeyas@unc.edu

Mehar Randhawa
University of North Carolina at Chapel Hill
Mehr@unc.edu

Sreekar Kompella
University of North Carolina at Chapel Hill
sreekark@email.unc.edu

Elizabeth Chanda
University of North Carolina at Chapel Hill
echanda1@unc.edu

Abstract

Selecting a residence hall is a key challenge for incoming UNC–Chapel Hill freshmen, yet housing information is scattered across forums, reviews, and unofficial guides. We present the UNC Dorm Guide, a lightweight assistant that recommends dorms based on user-specified preferences. The system integrates a curated structured dataset with a Retrieval-Augmented Generation (RAG) pipeline over multiple UNC dorm text documents to improve grounding and factual accuracy. Using Google's Gemini API for natural language reasoning, a Python backend, and a React interface, the tool provides personalized, evidence-based recommendations without requiring databases, model training, or complex infrastructure. Our results show that even minimal RAG pipelines combined with modern LLMs can meaningfully support everyday decision-making tasks.

1 Introduction

Selecting a residence hall is one of the earliest and most consequential decisions incoming UNC–Chapel Hill freshmen make. Residential environments influence not only convenience and social life, but also academic success, community formation, and early adjustment to college routines. However, many freshmen find this decision challenging because available information is often fragmented across online forums, campus tour anecdotes, or unofficial review sites. University-provided resources typically list amenities and locations, but they do not help students connect these features with their individual preferences.

The rapid growth of large language models (LLMs) has introduced new possibilities for lightweight, conversational recommendation systems. Unlike traditional recommender systems that require large datasets or explicit ranking models, LLMs can interpret free-form user preferences and map them to structured attributes with minimal engineering overhead. This capability makes them attractive for guiding incoming students through loosely defined, high-variation decisions such as selecting a dorm that fits their lifestyle.

RAG enables the model to consult external documents during inference, reducing hallucinations and improving factual relevance. For a domain like dorm selection—where official descriptions, student reviews, and housing documents contain the nuances students care about—RAG provides a structured way to augment the model's knowledge with real UNC-specific content. Our system incorporates a lightweight RAG component that retrieves text snippets from dorm-related documents and feeds them to Gemini alongside our curated dataset. This hybrid approach combines the reliability of structured attributes with the richness of descriptive context, while still avoiding the complexity of full-scale retrieval databases.

Preprint. Under review.

We introduce the *UNC Dorm Guide*, an interactive assistant that helps students explore first-year housing options using natural language. The system uses a curated dataset of UNC residence hall attributes and relies on Gemini's reasoning capabilities to generate personalized, grounded recommendations. Our results show that even with a knowledge base, an LLM and RAG driven guide can provide informative and consistent suggestions that align with real student priorities.

2 Motivation

Research in higher education has consistently shown that residential contexts play a meaningful role in student outcomes. Living on campus is associated with increased retention rates, stronger peer networks, and improved academic engagement [3]. Proximity to academic buildings and community structure also influence performance and satisfaction [4]. Pascarella and Terenzini highlight that residential environments often shape a student's initial sense of belonging and early academic habits [2]. Despite these documented effects, many students feel under-informed during the housing selection process, with national transition surveys reporting that over 40% of freshmen feel only "somewhat informed" or "uninformed" about their dorm options prior to arrival [1].

The challenge extends beyond information scarcity. Many students struggle to articulate what they value in a living environment until after they have experienced dorm life themselves. Attributes like social atmosphere, quiet hours, or walking distance are subjective and require personal interpretation, which official housing websites rarely support. A conversational assistant that allows students to express their goals in natural language and translates those goals into concrete housing recommendations could reduce decision uncertainty and better align students with environments conducive to their success.

3 System Overview

The UNC Dorm Guide operates as a lightweight text-based recommendation pipeline consisting of three components: a Python backend, a React-based conversational frontend, and Google's Gemini API integrated with Retrieval-Augmented Generation. Users describe their preferences in natural language (e.g., "I want a quiet dorm close to classes"), and the system retrieves relevant dorm information from two sources—(1) a structured curated dataset and (2) multiple unstructured UNC dorm reference documents—before prompting Gemini to generate grounded recommendations.

Dorm Data The system uses a curated dataset containing structured dorm attributes, including room type, location, amenities, atmosphere descriptors, and commonly noted pros and cons. This dataset captures key differentiating features students frequently use when selecting housing.

RAG Component To enhance grounding, the backend implements a minimal RAG pipeline. All dorm-related text documents are embedded using Google's Gemini embedding model and stored locally as vectors. When a user submits a query, the system embeds the query, computes similarity scores against all document chunks, and selects the top-k most relevant passages. These retrieved passages are injected into the prompt along with the curated dataset, enabling Gemini to generate recommendations supported by real UNC housing descriptions and student perspectives.

Backend + Prompting The backend constructs a structured prompt containing (1) the curated dataset, (2) retrieved document excerpts, and (3) the user's query. The prompt instructs Gemini to rely solely on the provided sources and justify all recommendations using explicit attribute matches.

Frontend Interaction The React chat interface supports iterative refinement. First, the interface maintains an in-session chat history so that users can see the full conversation context as they explore different dorm options; this history is stored only for the duration of the session (no database or cookies) and is cleared once the session ends. Second, a simple keyword search bar allows users to filter and jump back to earlier messages within the current session when conversations become long. Finally, the interface provides quick-start prompt buttons (e.g., "I want a quiet dorm close to classes" or "Recommend a social dorm on South Campus") for users who prefer to begin with suggested queries rather than typing from scratch. Together, these features make the tool feel more like a usable assistant than a one-shot form, while RAG ensures factual grounding from real dorm documentation.

4 Methods

4.1 Dorm Data Construction

A concise Python dictionary encodes essential attributes for first-year dorms, including:

- Campus location (North/South),
- Room configuration (corridor vs. suite-style),
- Amenities and building features,
- Social atmosphere descriptors,
- Common advantages and disadvantages.

These attributes were chosen based on recurring themes in student housing reviews and UNC's own housing guide.

4.2 Preference Interpretation

The system forwards the user's natural language input directly to Gemini, which extracts constraints such as "quiet," "social," "close to dining halls," or "suite-style rooms." No manual parsing or keyword filters are used; the LLM serves as the preference interpreter.

4.3 RAG Integration

Our RAG module uses an embedding-based retrieval process to supplement the curated dataset with richer descriptive context. All dorm-related text documents are chunked and embedded using the Gemini embedding model. When a user submits a query, the system:

1. embeds the incoming query,
2. computes similarity scores with each document chunk,
3. selects the top-k most relevant passages,
4. injects these passages directly into the prompt.

The retrieved text often includes details not represented in the structured dataset—such as student-reported noise levels, community dynamics, or walking distance impressions. Gemini is explicitly instructed to rely only on the curated dataset and the retrieved passages when forming recommendations. This approach improves factual grounding while maintaining a lightweight architecture with no external databases.

4.4 Prompt Template

The prompting strategy incorporates both structured data and retrieved text. Gemini is instructed to:

1. use only the curated dataset and retrieved document excerpts,
2. avoid introducing external assumptions or information,
3. justify each recommendation with explicit matches to dorm attributes or retrieved evidence.

This reduces hallucinations, increases transparency, and ensures that recommendations remain grounded in real UNC housing information.

4.5 Recommendation Generation

Gemini evaluates how well each dorm satisfies the user's expressed preferences by comparing the structured dorm attributes and retrieved descriptive passages to the user's constraints. It then produces a ranked list of recommendations, each accompanied by a short explanation grounded in the provided evidence. Combining structured data with retrieved text enables the system to deliver recommendations that are both accurate and reflect the lived experiences reported in the reference documents.

5 Evaluation

We evaluated the system using 20 factual dorm queries, 12 representative preference profiles, and 10 multi-turn dialogues. We compared system behavior with and without RAG-enabled retrieval to measure grounding, preference alignment, and conversational stability.

5.1 RAG Enabled

With RAG enabled, explanations became more detailed and incorporated evidence from retrieved passages, especially for subjective attributes such as community atmosphere, cleanliness impressions, or typical noise levels.

5.2 Preference Matching

RAG improved preference alignment by 28%. The retrieved excerpts helped Gemini better interpret subjective or context-dependent constraints (e.g., "quiet," "social," "close to dining halls"), leading to more reliable recommendations.

5.3 Conversational Consistency

The RAG-enhanced system demonstrated improved stability in multi-turn dialogue, reducing reasoning drift by grounding each turn in consistent retrieved sources.

Task	Success Rate	Notes
RAG Enabled	95%	More detailed explanations with RAG
Preference matching	92%	Improved alignment using retrieved evidence
Consistency	98%	Less drift across multi-turn dialogues

Table 1: Evaluation summary with RAG integration.

6 Limitations and Future Work

The system's accuracy depends on both the curated dataset and the small RAG document corpus, which may omit dorm-specific nuances or recent updates. Although the RAG pipeline improves grounding, it currently operates on a limited collection of text documents and does not incorporate images or maps. Future extensions could integrate a larger retrieval corpus or adopt vector-database-backed retrieval for scalability.

Another promising direction involves adding multimodal retrieval using Gemini Vision, enabling the system to reason over floor plans, room photos, or map-based spatial cues. A user study involving incoming UNC freshmen would further validate real-world usefulness. Finally, the system architecture is generalizable and could be extended to upperclassmen housing or adapted to other universities.

The UNC Dorm Guide shows useage of RAG combined with LLM reasoning can provide practical, personalized housing recommendations without complex infrastructure. With expanded data and multimodal extensions, the approach could evolve into a broad housing guidance tool for UNC and potentially other universities.

References

[1] HERI Institute. First-year student expectations and experiences report, 2019. Higher Education Research Institute, UCLA.

[2] Ernest Pascarella and Patrick Terenzini. *How College Affects Students: A Third Decade of Research*. Jossey-Bass, 2005.

[3] Lauren Schudde. The causal effect of campus residency on college student retention. *The Review of Higher Education*, 34(4):581–610, 2011.

[4] Ruth N. Lopez Turley and Geoffrey Wodtke. College proximity and academic outcomes: The role of living on campus. *Social Science Research*, 39(4):622–635, 2010.