Title	**Lab Link**
Course	CS/QTM/LING-329: Computational Linguistics
Authors	Emma Carrier, BS in QSS-Data Science, `emma.carrier@emory.edu`
	Maya Dattatreya, BS in Neuroscience and Behavioral Biology, `maya.dattatreya@emory.edu`
	Rhea Kansal, BS in Neuroscience and Behavioral Biology, `rhea.kansal@emory.edu`
	Ryan Lin, BS in CS and Math, `ryan.lin@emory.edu`
	Alaş Özen Sezgin, BS in Neuroscience and Behavioral Biology, `alas.ozen.sezgin@emory.edu`

Abstract

Lab Link is a conversational chatbot designed to help Emory University students discover research labs that align with their academic interests, skills, and goals. Unlike traditional methods that rely on outdated lab websites, Lab Link uses real-time publication data from sources like Google Scholar using the Scholarly Python package to provide up-to-date insights into ongoing faculty research within the Computer Science (CS) and Quantitative Theory and Methods (QTM) departments. Using natural language processing (NLP) and Google's Gemma model, we developed an LLM based approach to analyze recent faculty publications, allowing students to engage with current research rather than static summaries.

The primary goal of Lab Link is to streamline the research discovery process through a web-based chatbot that enables students to describe their interests conversationally and receive personalized, actionable recommendations. By the end of the semester, we aim to deliver a working chatbot interface and a backend pipeline to collect and analyze faculty publications that helps students discover research labs.

The technical scope of this phase is limited to faculty in the CS and QTM departments. While students will interact with the system through natural language input, real-time chat with advisors or faculty is outside the current scope.

1 Introduction

1.1 Objectives

The primary objective of Lab Link is to develop a user-friendly and accessible platform that connects Emory University students with research opportunities tailored to their academic interests and skillsets. By addressing the existing gaps in how research positions are discovered and accessed, Lab Link aims to streamline and personalize the process, making research engagement more inclusive and efficient.

To achieve this goal, one of our core objectives is to develop a conversational chatbot interface that allows students to naturally describe their academic interests, goals, and experiences. This interface will eliminate the need for technical jargon or predefined keywords, making the platform approachable for students at all levels of research experience.

Another key component is the creation of a research data collection pipeline. By leveraging sources such as Google Scholar and the Scholarly Python package, Lab Link will gather up-to-date information on faculty publications, ensuring that recommendations are grounded in current research activity rather than outdated website descriptions [1].

To use this data, we will leverage Google's Gemma model to assist students in matching with their ideal labs. We will finetune the pretrained Gemma model on our own custom dataset to allow the model to make more informed decisions based on specific and recent data from the Emory research community.

Finally, Lab Link will include a recommendation system that presents students with personalized, actionable research opportunities in a clear and intuitive format. This system will empower students to make informed decisions and confidently pursue positions that align with their academic passions and long-term goals.

1.2 Motivation

At Emory, students often struggle to find research labs that align with their specific academic interests, despite the wide range of research opportunities available across departments [7]. This disconnect between students and faculty-led research is not due to a lack of opportunities, but rather a series of structural and informational barriers that make it difficult for students to identify and pursue relevant positions [2].

One key issue is that many lab websites are outdated and fail to reflect the current research activities, priorities, or availability of positions. Without accurate, up-to-date information, students are left guessing which labs may be a good fit for their interests and skills. Additionally, there is no centralized system at Emory that helps match student interests with faculty research areas, leaving students to conduct independent searches that are often overwhelming and inefficient.

The process of researching each lab individually can be extremely time-consuming, especially for students juggling coursework, extracurriculars, and part-time jobs. As a result, many students miss out on valuable opportunities—not because they lack interest or ability, but because of informational and logistical barriers.

This project is motivated by the need to close that gap. Lab Link seeks to create a more efficient, accurate, and equitable system for connecting students with faculty research. By streamlining access to current research information and matching students with labs based on their individual interests and skills, we aim to make undergraduate research at Emory more accessible, inclusive, and impactful.

1.3 Problem Statement

The core problem Lab Link aims to solve is the inefficiency and information asymmetry in the current research lab discovery process at Emory University. Many students struggle to find research opportunities that align with their academic interests, goals, and skill sets. This is largely due to the lack of a centralized, user-friendly system that provides up-to-date information about ongoing faculty research.

These challenges not only create frustration for students eager to gain research experience, but also result in missed opportunities for faculty who are seeking motivated, capable research assistants. By addressing these inefficiencies, Lab Link seeks to bridge the gap between students and faculty, making the research matching process more transparent, accessible, and effective for all involved.

1.4 Innovation

Lab Link introduces several innovative approaches to address the challenges students face in discovering research opportunities. One key innovation is the use of recent faculty publications—specifically from 2019 to 2025—as the primary source of information, rather than relying on potentially outdated lab websites. This ensures that students are matched with faculty who are actively conducting research in their stated areas of interest, providing a more accurate and timely picture of the research landscape at Emory.

Another core feature of Lab Link is its conversational chatbot interface, which makes the discovery process more accessible and intuitive. Instead of requiring students to navigate complex databases or know specific terminology, the chatbot allows them to describe their interests in natural language. This lowers the barrier to entry and makes it easier for students from diverse academic backgrounds to engage with research opportunities.

Additionally, Lab Link shifts the focus of research matching from historical lab descriptions to current research activity. By analyzing recent publications, the system is better equipped to identify emerging topics and align students with cutting-edge projects. Finally, Lab Link is designed to be a self-improving platform. It continuously learns from user interactions and feedback, refining its recommendations over time to better serve both students and faculty.

2 Background

Several existing systems attempt to connect students with research opportunities, but each comes with significant limitations. University department websites, for instance, often maintain faculty directories that list general research interests. However, these pages are frequently outdated and rarely provide specific or current information about ongoing projects, making it difficult for students to assess whether a lab aligns with their interests.

Research opportunity databases like ForagerONE offer some centralized listings, but they typically rely on faculty to manually submit opportunities [4]. As a result, many active research labs are never listed. Additionally, programs such as Emory's SIRE (Scholarly Inquiry and Research) are more accessible to students who have already identified a professor to work with, rather than those who are still exploring potential matches.

General career services platforms used by universities may include research listings, but they often lack the granularity and academic focus necessary for meaningful research alignment. These platforms are designed for broad job or internship searches and are not optimized for the specific needs of undergraduate research matching.

Manual advising also plays a role in connecting students with labs. Academic advisors and faculty members can recommend labs based on a student's interests, but this approach is not scalable and heavily depends on the advisor's personal familiarity with current research across departments. Lastly, many students rely on word-of-mouth networks to find research positions, which creates disparities for those without established academic or social connections—particularly first-generation college students or those new to the research ecosystem.

Lab Link addresses these limitations by offering a centralized, intelligent, and up-to-date system that levels the playing field for all students seeking research opportunities.

3 Methodology

3.1 Data Collection and Processing

To build an accurate and up-to-date picture of ongoing research in the Emory CS and QTM departments, we will use the Scholarly Python package [3] to collect Google Scholar publications within a 5-year period (2019-2025) from faculty. This Python package allows us to gather important information about each publication a certain professor has published, which will be used to inform our chatbot in future steps.

To begin the generation of our dataset, we will first gather a list of all professors in the respective departments. Scholarly allows us to search for publications based on author name, so we can utilize our list of all faculty to begin generating a database of all publications. For each publication, we would store the title, author / coauthors, date published, URL, and abstract. This data set will provide us with the centralized information needed to create a reliable chatbot that can find relevant research lab opportunities for students.

The data collected from the publications will be then used to generate a JSON file with input-output pairs for finetuning the LLM.

3.2 Chatbot Development

The next step to develop Lab Link is to create a chatbot. This chatbot will answer questions based on the custom data set that we previously described. To implement the chatbot functionality, we plan to integrate Google's Gemma, an open-source LLM designed for building custom AI applications. Google already has a pre-trained Gemma model available to the public, which is what we will be using for this project.

To make our chatbot preform better than other LLMs at this task, we will finetune the Gemma model with the data from the CS and QTM departments. To finetune the model, we will utilize LoRA (low-rank adaption) tuning as outlined in Google's official documentation [5]. The tuning process uses the gemma-transformers library in combination with the Hugging Face ecosystem. A pretrained Gemma model is first loaded using the transformers interface, and the peft (Parameter-Efficient Fine-Tuning) library is used to apply LoRA adapters to specific parts of the model, such as attention layers. Only these adapter parameters are trained, while the rest of the model remains frozen. To perform the tuning, we will provide our previously generated JSON file and define a training configuration that includes LoRA specific-settings.

For each match, the chatbot will explain why the faculty member was suggested, referencing specific publications and research areas to provide context. Additionally, the chatbot will have the ability to make corrections or changes to the recommendations provided using feedback. This can be in the form of an additional message from the user asking to add an interest of theirs, or to narrow down the scope of their interests even further. Many LLMs do not have built-in memory, which means we will need to provide the entire conversation history as input for each new message. This functionality will allow for feedback and changes to be made from the original prompt sent in by the user.

3.2.1 Future Functionality

Once the framework for this chatbot is working efficiently on our self-hosted database, we can consider implementing RAG (Retrieval-Augmented Generation) functionality [6] The benefit of implementing RAG would be that the chatbot can access external information, in this case our database of research articles. Future steps for this project could include creating a database that is automatically updated when a new research paper is published by an Emory professor, and the RAG functionality of the chatbot will allow these research papers to be considered in the matching process.

3.3 UI Creation

The user interface for Lab Link will be developed using a modern web stack to ensure accessibility and responsiveness across various devices. The front-end implementation will utilize React.js for component-based architecture, complemented by Tailwind CSS for streamlined styling and responsive design. The interface will feature a conversational chat window as its central element, allowing students to input their research interests, academic background, and skill sets through natural language interaction. To enhance usability, the system will incorporate guided conversation prompts and suggestion buttons that facilitate information input without requiring extensive typing. The UI will present faculty matches in card-based formats with expandable sections that reveal publication details, research summaries, and relevance explanations.

4 Evaluation Metrics

To ensure the effectiveness of Lab Link, we will evaluate performance using mostly human-centered feedback. To evaluate the effectiveness and user experience of the matching and ranking system, we can survey participants that agree to test out Lab Link. After explaining to the users the purpose and functionality of the tool, they can utilize Lab Link to find a few research labs that align with their interests and identify one they would consider applying to. Based on their experience with the matching system and chatbot performance, users will respond to a few questions. Some sample questions include: "The model helped me find relevant research opportunities," "The chatbot accurately understood my research interests," and "I would prefer this tool over other AI companions I've used." Each question used a 5-point Likert scale, supplemented by open-ended prompts to gather qualitative insights. This feedback can be used to adjust our model or chatbot accordingly to better align with the needs of Emory students.

Although out of scope for this course, an additional evaluation that could be done for this project is to conduct follow-up surveys with students who have used Lab Link to find labs they are interested in. Some metrics that can be taken from this would be if the students actually applied to the research labs they were suggested, if they got the opportunity to join the lab, etc. Feedback like this could inform future changes to the model and implementation on a large scale.

5 Timeline

5.1 Weekly Schedule

- **Week 1 (March 17-24)**: Project setup and initial requirements gathering

- **Week 2 (March 24-30)**: Data collection system and publication processing pipeline
- **Week 3 (March 31-April 6)**: Gemma training and chatbot foundation
- **Week 4 (April 7-13)**: User interface development and system integration
- **Week 5 (April 14-21)**: Testing, refinement, and demo preparation
- **Live Demonstration (April 23)**
- **Week 6 (April 24-30)**: Address feedback and implement improvements
- **Week 7 (May 1-5)**: Final testing and report preparation
- **Final Report Due (May 5)**

5.2 Team Responsibilities

Alaş:

- Develop and implement the publication data collection system
- Create the processing pipeline for research papers
- Manage database structure and data storage

Rhea:

- Implement conversational chatbot functionality
- Create input-output responses for chatbot training from data
- Gemma model assistance

Ryan:

- Design and implement the web interface
- Create responsive design for multiple devices
- Develop user experience flow

Emma:

- Implement conversational chatbot functionality
- Develop a finetuned Gemma model
- Create response generation system

Maya:

- Coordinate testing efforts
- Collect and analyze user feedback
- Manage documentation and reporting
- Ensure quality assurance across all components

5.3 Progress Tracking Plan

To ensure timely completion and alignment with project goals, a comprehensive progress tracking system will be implemented. Each week begins with defining specific, measurable objectives tied to timeline milestones, with progress measured through a project dashboard displaying completion percentages for major deliverables. Bi-weekly retrospectives will evaluate process effectiveness and identify improvement opportunities. Bi-weekly reviews will compare actual progress against the project roadmap, with adjustments made as needed. This structured approach ensures transparency and accountability while maintaining flexibility to adapt to emerging challenges.

5.4 Collaboration Methods

The team will employ a structured yet flexible collaboration framework to maximize productivity while maintaining quality. Bi-weekly synchronous meetings for progress updates and blockers discussion will be implemented, complemented by daily asynchronous communication through iMessage for immediate questions and updates. Code collaboration will utilize GitHub with a branch-based workflow, requiring peer review before merging and maintaining comprehensive documentation. Shared documentation using Notion will be established to centralize project requirements, technical decisions, and progress tracking.

5.5 Risk Mitigation Strategies

The project faces several potential challenges including data availability limitations, algorithm accuracy issues, and user adoption barriers. To mitigate these risks, a multi-layered approach will be implemented. For data collection risks, backup data sources including direct faculty input will be established if API constraints or publication accessibility issues arise. For user experience risks, early prototype testing with a sample student group will be conducted to identify usability issues before full deployment. By anticipating these challenges proactively, the project can progress even if primary methods encounter obstacles.

References

[1] Yusuff Adebayo Adebisi. Undergraduate students' involvement in research: Values, benefits, barriers and recommendations. Annals of Medicine and Surgery, Vol. 81, 104384, 2022. Accessed: 2025-03-24.

[2] Angela Brew and Lilia Mantai. Academics' perceptions of the challenges and barriers to implementing research-based experiences for undergraduates. Teaching in Higher Education, Vol. 22, No. 5, pp. 551-568, 2017. Accessed: 2025-03-24.

[3] Scholarly Contributors. Scholarly: A Python Scraper for Google Scholar. Python Package, 2024. URL `https://pypi.org/project/scholarly/`. Accessed on March 24, 2025.

[4] Frederick Daso. Foragerone lays the foundation for university research initiatives, 2020. URL `https://www.forbes.com/sites/frederickdaso/2020/04/12/foragerone-lays-the-foundation-for-university-research-initiatives/`. Accessed: 2025-03-24.

[5] Google AI for Developers. Fine-tune gemma models in keras using lora, 2025. URL `https://ai.google.dev/gemma/docs/core/lora_tuning`. Accessed: 2025-03-24.

[6] Mandar Kulkarni et al. Reinforcement learning for optimizing rag for domain chatbots. arXiv preprint arXiv:2401.06800, 2024. Accessed: 2025-03-24.

[7] Lennon Tomaselli. Opinion: Students shouldn't struggle to find lab accessibility, 2023. URL `https://www.usforacle.com/2023/10/12/opinion-students-shouldnt-struggle-to-find-lab-accessibility/`. Accessed: 2025-03-24.

A Contributions

- Ryan Lin - 20
- Alaş Özen Sezgin - 20
- Emma Carrier - 20
- Rhea Kansal - 20
- Maya Dattatreya - 20