GeoRisk AI

An Agentic Geospatial AI Chatbot for Autonomous Multi-Domain Urban Risk Analysis

Columbia University, Data Science Institute

Fall 2025

Rishi Kumar (rk3326), Louis Chu (lc3920), Sharon Tsai (ct3263), Max Jan B Sebrechts (mjs2447), Kevin Ke (yk3108)

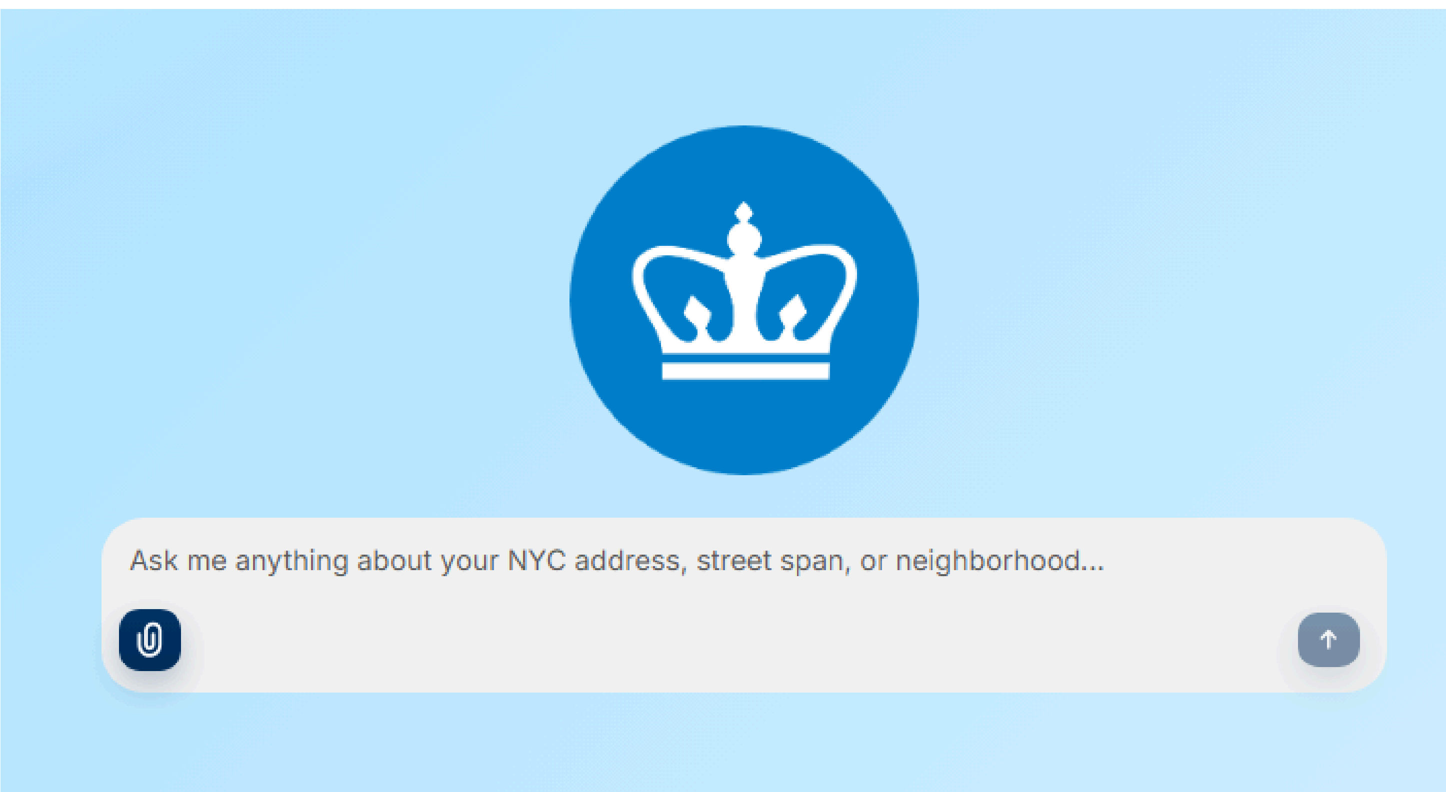


Table of Contents

1. Problem Description

This Capstone project is a collaboration between **Columbia University** and **Town+Gown: NYC @ DDC Project Controls.**

1.1 Motivation and Background

Public infrastructure and building projects in New York City must comply with a variety of environmental, zoning, and safety regulations. However, the information needed to assess project risk is scattered across dozens of publicly available datasets maintained by different agencies—each with its own formats, geospatial units, and update schedules.

Project managers at the NYC Department of Design and Construction (DDC) currently spend significant time manually cross-referencing data from sources such as MAPPLUTO, LION, environmental datasets, crime datasets, and many others to evaluate whether a proposed project site lies within areas of concern (e.g., asbestos contamination, flood risk, etc.). This manual, spreadsheet-driven process slows down project reviews and increases the likelihood of human error, especially when multiple datasets use inconsistent geospatial reference systems.

1.2 Problem Statement

The core problem this project addresses is **the fragmentation and complexity of geospatial risk data used by NYC project managers**. Each dataset is indexed by different geounits, such as Borough-Block-Lot (BBL), Building Identification Number (BIN), Street Segment (LION), Neighborhood Tabulation Area (NTA), and Precinct, among others, making it challenging to automatically align information across data sources to relevant project sites.

Our goal is to develop **GeoRisk AI, an autonomous, agentic geospatial AI chatbot** that interprets natural-language user queries (e.g., questions about an address), converts them into appropriate geospatial units, dynamically selects and retrieves relevant risk datasets, and synthesizes the results into clear, project-manager-friendly insights. The system operates through a self-directed, agentic workflow that adapts its actions at each step while maintaining an ongoing, multi-turn conversation—all without requiring users to have prior geospatial expertise.

1.3 Project Overview

Recent research has highlighted the growing potential of chatbots and large language models (LLMs) for context-aware reasoning. *Zhu et al.* [1] showed that chatbot-based hazard training can improve safety awareness in construction settings, while *Polo-Rodríguez et al.* [2] demonstrated how LLM-driven systems can leverage location context to enable adaptive, real-time interactions.

However, more recently, research on **agentic AI systems** has examined how generative models can autonomously decompose high-level user goals into multi-step actions, interact with external tools, and provide transparent reasoning traces. *Brachman et al.* [4] studied user mental models and trust in an agentic AI chatbot capable of selecting actions and information sources independently, highlighting both the power and design challenges of such systems. Building on these advances, GeoRisk AI applies agentic AI principles to the urban infrastructure domain, enabling autonomous geospatial reasoning and dataset selection while presenting results in a form accessible to non-expert project managers.

Thus, GeoRisk AI functions as an **agentic geospatial intelligence assistant**, enabling project managers and city officials to pose natural-language questions such as *"What environmental risks are near 237 Park Avenue?"* Rather than following a fixed execution path, the system **autonomously plans and executes** a sequence of geospatial and data operations tailored to each query. Through this agentic data-to-dialogue workflow, GeoRisk AI produces concise, evidence-based summaries while maintaining conversational continuity across turns.

At the core of the system is an **agentic controller** that determines *what to do next* at each stage—whether to parse additional location details, expand spatial scope, retrieve new datasets, reuse previously fetched data, or generate follow-up questions. This design enables automation, interpretability, and scalability without requiring users to have prior expertise in geospatial analysis.

1. **Data Ingestion & Preprocessing:** GeoRisk AI integrates datasets spanning multiple urban risk domains. All datasets are standardized to a common **Borough–Block–Lot (BBL)** representation using geometric joins, the **NYC Geoclient API**, and conversion adapters. This unified spatial backbone enables consistent querying and seamless interoperability across heterogeneous data sources.

2. **Geospatial Normalization & GeoScope Generation:** When a user provides a location, the system retrieves canonical identifiers (BBL, BIN, latitude/longitude) via the Geoclient API. Rather than relying on a fixed spatial buffer, the **GeoScope module is invoked agentically**, allowing the system to decide whether to analyze only the target lot or expand to surrounding parcels, blocks, or street spans using **LION street geometry**. This enables contextual risk assessment that adapts to the user's intent.

3. **Agentic Query Interpretation & Dataset Selection:** An **LLM-based parser** interprets user intent, extracts structured location information, and identifies one or more relevant risk domains (e.g., Environmental & Health Risks, Construction & Permitting). Crucially, the agent **autonomously selects which datasets to query** using a configurable routing dictionary and may expand or refine its selection as the conversation evolves, ensuring efficient retrieval and coherent downstream reasoning.

4. **Risk Retrieval & Summarization:** The **DataHandler** retrieves and filters datasets based on the agent-defined GeoScope, joining records into a unified analytical view. A **specialized summarization LLM** then synthesizes this information into concise, natural-language risk summaries—distinguishing between active, historical, and upcoming risks—while grounding all outputs in retrieved data.

5. **Conversational Control & Iterative Dialogue:** A **conversational LLM** manages dialogue state and memory, enabling agent-driven follow-ups such as *"include nearby lots," "show the underlying records,"* or *"compare with another address."* The agent decides whether a follow-up requires new data retrieval or can be answered from existing context, supporting efficient multi-turn interaction and exploration.

Together, these components form an **end-to-end agentic architecture** that transforms unstructured user queries into structured geospatial intelligence and, in turn, into clear, context-aware responses. By combining deterministic geospatial computation (e.g., address normalization, GeoScope generation, spatial joins) with generative reasoning (LLM-based interpretation, planning, and summarization), GeoRisk AI delivers outputs that are both **factually grounded and adaptively conversational**. This architecture establishes a scalable foundation for integrating additional datasets and extending the system to broader multi-domain urban risk assessment.

2. Methods

2.1 Primary Data Sources

All datasets used in this project were sourced from NYC Open Data, the City of New York's official website for providing public information across domains such as zoning, infrastructure, and environment. Of the 20+ datasets identified (see *Appendix A1* for further details on each dataset), MapPLUTO and LION form the project's geospatial framework, while the remaining datasets serve as domain-specific data for risk insights. MapPLUTO provides Borough-Block-Lot (BBL), a unique identifier used in NYC to locate and reference individual tax lots that serve as the primary spatial key. LION contains street segment geometries that link intersection-based queries to nearby BBLs. The NYC Geoclient API complements them by geocoding addresses and returning related BBL and street information. Because datasets reference different geounits (e.g., BINs, NTAs, precincts, coordinates), all were standardized to the BBL level using MapPLUTO. Additional sources, such as LION and NTA2020, were required to map street spans and Neighborhood Tabulation Areas, which Geoclient does not provide. This ensured consistent spatial alignment before integration into the chatbot.

2.1.1 Primary Land Use Tax Lot Output - Map (MapPLUTO) and LION

The MapPLUTO dataset, maintained by the NYC Department of City Planning, combines detailed tax lot information with land-use, zoning, and building data. It provides parcel-level geometries across the five boroughs of NYC using BBL alongside other geospatial identifiers. MapPLUTO served as one of the primary reference layers, as it contains BBL polygons that could be overlapped with geometries from all other datasets. To optimize performance, the dataset was loaded using Python's lru_cache decorator, which enabled efficient memory management while maintaining fast access to lot geometries. The LION dataset shared similar functionality to the MapPLUTO dataset. Although it doesn't include a BBL identifier, it covers public infrastructure, such as roads and utilities, using street and borough codes. In this project, LION was used to link street-level datasets (e.g., construction and traffic data) to their corresponding BBLs by matching spans and intersections. The dataset was read and processed similarly, using lru_cache to maintain the same memory-efficient strategy.

2.2 Geoclient API

Once the datasets are standardized using MapPLUTO and LION, the NYC Geoclient API was integrated to handle natural-language user queries. When a user enters a location prompt (e.g., "Summarize the asbestos risk in 237 Park Ave."), the LLM-parsed input is fed to the Geoclient API to retrieve attributes such as BBL, BINs, longitude, and latitude. The retrieved BBL will then be merged and used to search across all datasets for relevant information. This approach ensured that all datasets, regardless of their original geospatial unit, can be queried and integrated consistently through a unified geospatial framework.

3. Exploratory Data Analysis

3.1 Risk Category Definition

During the exploratory phase, our primary objective was to understand the scope, spatial resolution, and thematic relevance of available NYC Open Data sources. After examining over 20 datasets, we organized them into six major **risk categories—Environmental & Health Risks, Zoning & Land Use, Construction & Permitting, Transportation & Traffic, Public Safety & Social Context, and Comparative Site Queries** (see details in *Appendix A4*). These categories capture the main types of questions city agencies and project managers are likely to encounter.

Each dataset may belong to multiple categories. For example, the *Asbestos Control Program* dataset contributes to both **Environmental & Health Risks** and **Construction & Permitting**, since asbestos filings are both environmental and construction-related. This tagging process produced a configuration mapping (cat_to_ds) that now enables the LLM query router to automatically match user intent (e.g., "environmental risks") to the correct datasets.

3.2 Geounit Exploration and Standardization

In the previous section, we discussed that different NYC Open Data sources describe location using different geospatial units (geounits). To address this challenge, we systematically examined the geounits in each dataset and developed a conversion framework to unify them into a standard reference unit.

3.2.1 Choose BBL as the Common Spatial Unit

The **Borough–Block–Lot (BBL)** is the most fundamental parcel identifier used across New York City's property, zoning, and environmental datasets. We selected BBL as our *base geounit* because it offers three key advantages. First, it is widely used across critical datasets, including MAPPLUTO, Department of Buildings (DOB) permits, and zoning maps, making it a universal reference point for geospatial alignment. Second, it provides fine-grained parcel-level precision, ideal for construction and land-use risk analysis. Finally, other spatial identifiers—such as BINs, precincts, or neighborhood tabulation areas—can be reliably mapped to BBL through geometric or API-based transformations (e.g., the NYC Geoclient API or polygon joins). By unifying all datasets at the BBL level, we ensured that data from different sources could be cross-joined, aggregated, and visualized consistently in downstream analyses.

3.2.2 Building Geounit Adapters

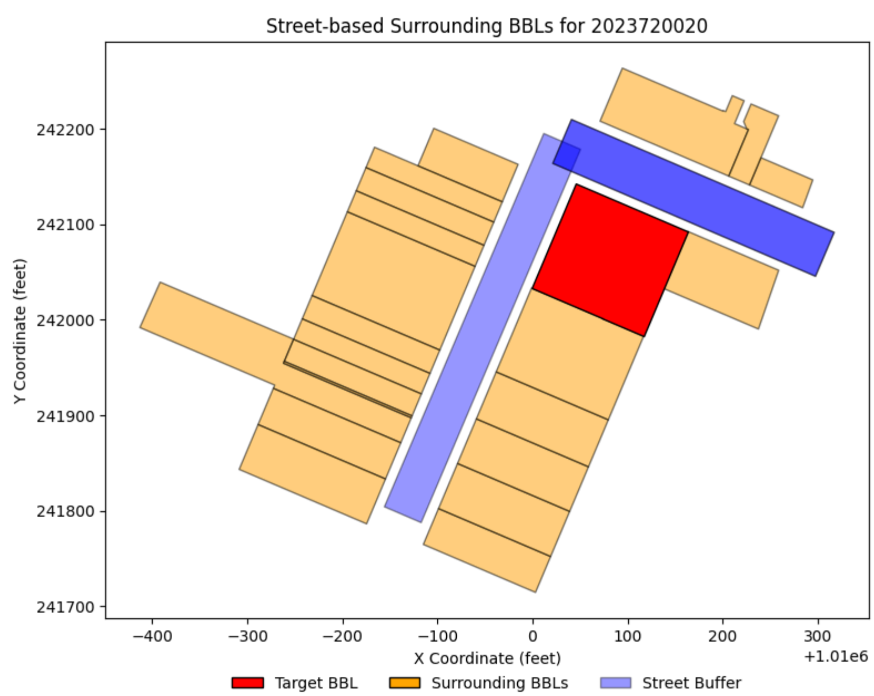
The key geounits we encountered and integrated are provided in *Appendix A3*: Geounits. To translate between these spatial references, we developed modular "adapter" scripts. Each adapter converts between BBL and another geounit type (bidirectional), allowing flexible filtering and merging across datasets. Each adapter can be called independently, but all follow a consistent API, allowing the system to automatically pick the correct conversion route based on a dataset's metadata (e.g., whether it uses NTA, BIN, or raw coordinates) and the agent's understanding of the user's intent.

3.2.3 From Geounits to GeoScope Mechanism

While BBL standardization allows all datasets to be compared at a parcel level, risk evaluation often requires analyzing context — not just a single lot, but also its surrounding environment. To achieve this, we implemented a *GeoScope mechanism* that automatically determines the spatial scope of each user query. GeoScope starts with a user's address (converted to BBL via Geoclient) and finds surrounding parcels via spatial buffering along the street network (using LION). When GeoScope determines a set of relevant BBLs for a user query, the system checks each dataset's configuration (GeoConfig) to identify what spatial unit it uses. Then, it automatically applies the corresponding adapter to translate the BBL-based scope into the dataset's native unit before applying filters.

Given a user input: *"Are there asbestos filings near 237 Park Avenue in Manhattan?"*, GeoScope performs three key steps:

1. **Address → BBL:** The system first calls the NYC Geoclient API to convert the address into a standardized BBL.
2. **BBL → Surrounding BBLs:** Using LION street geometry and MAPPLUTO, it finds all nearby parcels that are spatially related to the input BBL.
3. **Scope → Dataset Filter:** The surrounding BBLs are then translated into each dataset's native geounit (e.g., coordinates, precincts) using the appropriate *geounit adapter*. These converted spatial units are used to filter relevant records (e.g., asbestos filings, flood zones, or permits) before risk summarization.



4. Agentic LLM Integration and Workflow

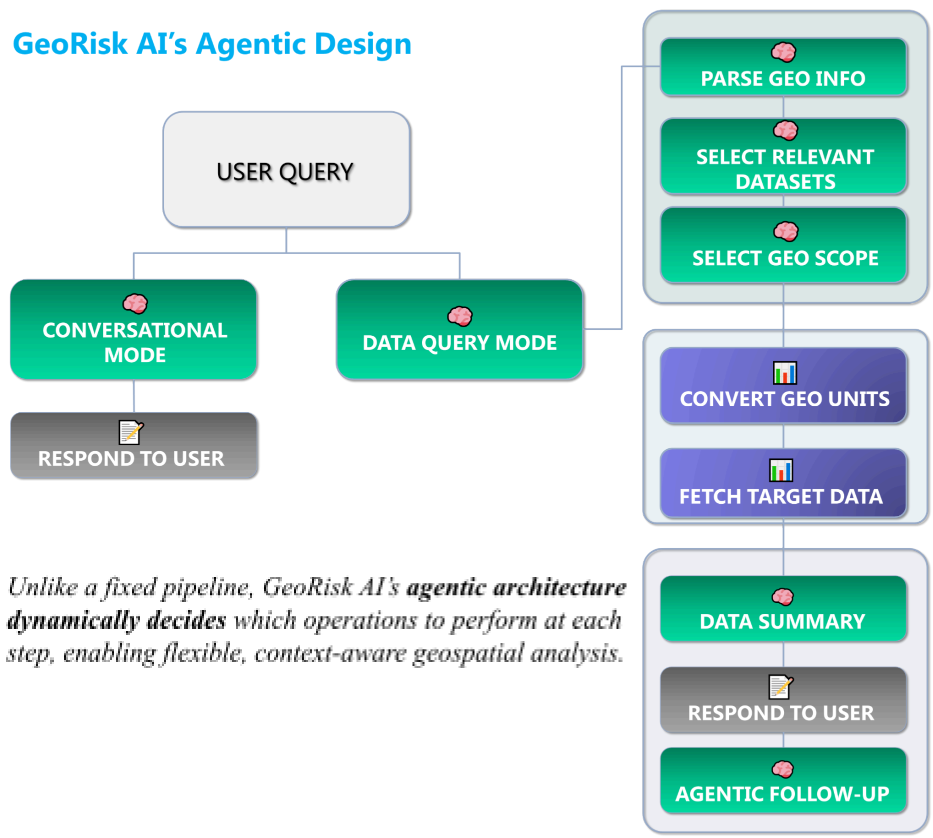
While incorporating deterministic geospatial standardization, as discussed in the previous section, GeoRisk AI also adopts an **agentic architecture** in which large language models are used not solely for text generation but also as **decision-making components** that autonomously control how the system processes each user query. Rather than executing a fixed pipeline, the application dynamically determines which actions to perform—such as parsing locations, selecting datasets, expanding spatial scope, retrieving data, or generating follow-up questions—based on the user's intent and conversational context.

At the center of this design is a **stateful conversational LLM**, which serves as the user-facing agent. This conversational agent maintains dialogue context across turns, decides how to respond at each step, and orchestrates interactions with downstream modules. In addition to this primary agent, the system employs several **stateless, task-specific LLM calls** (e.g., parsing, risk summarization) that are invoked conditionally as tools within the agent's workflow. All LLM interactions are performed via the Gemini API, supported by GCP credits provided for the course.

This agentic design was chosen to address two key challenges: (1) the variability and ambiguity of natural-language geospatial queries, and (2) the need for flexible, multi-step reasoning when working with heterogeneous urban risk datasets. By allowing the system to determine the next action rather than enforcing a rigid execution order, GeoRisk AI achieves greater adaptability, efficiency, and interpretability.

4.1 Agentic Query Processing and Control Flow

Each interaction begins with a user query submitted to the conversational LLM. Rather than immediately executing all downstream steps, the agent first determines the appropriate **mode of operation**, such as responding conversationally, initiating data retrieval, or requesting clarification. When data access is required, the agent invokes a varying sequence of task-specific components to support its reasoning.



*Unlike a fixed pipeline, GeoRisk AI's **agentic architecture dynamically decides** which operations to perform at each step, enabling flexible, context-aware geospatial analysis.*

An **LLM-based Parser** is used as an agent tool to extract structured information from unstructured user input. This includes identifying one or more addresses, classifying the query into relevant risk domains, and determining which datasets may be required. These decisions are not static; the agent may reuse previously extracted information, expand the set of datasets, or revise its interpretation based on conversation history.

The extracted address information is passed to the GeoUnit adapters and GeoScope module, which the agent invokes to normalize locations into canonical geospatial units (e.g., BBL, BIN, latitude/longitude). Importantly, the **agent**

determines whether to analyze only the target location or to expand the spatial context—for example, by including surrounding parcels, blocks, or street spans—based on the nature of the query.

Once the agent determines that sufficient data has been retrieved, it may choose to preview underlying records for transparency, proceed directly to summarization, or generate a follow-up prompt to guide further exploration. This flexible control flow enables efficient multi-turn interactions while avoiding unnecessary computation.

4.2 LLM Parser

The LLM Parser operates as a **decision-support tool within the agentic workflow**, enabling the system to translate unstructured user input into structured signals that guide downstream actions. It performs three core functions: address extraction, query classification, and dataset selection.

First, the parser extracts all addresses mentioned in the user query and returns them in a structured JSON format. This is achieved through a carefully engineered prompt refined via few-shot examples to improve robustness against ambiguous or incomplete inputs. If the LLM fails to produce a valid structured output, a fallback regular-expression-based extractor is used to ensure continuity. Second, the parser classifies the query into one or more predefined risk categories (as described in Section 3.1 and *Appendix A5*). For each selected category, the LLM produces a confidence score, following the approach of *Tian et al.* [3], which has been shown to improve calibration and downstream decision-making. These confidence scores allow the agent to prioritize or refine dataset selection. Finally, the selected categories are mapped to relevant datasets using a configurable routing dictionary. This enables the agent to retrieve only information pertinent to the query, rather than exhaustively querying all available datasets, and to adapt its selections dynamically as the conversation evolves.

4.3 LLM Risk Summarization

The LLM Risk Summarization module **synthesizes retrieved geospatial data** into concise, actionable insights aligned with the user's query. Rather than operating in isolation, this module is invoked **agentically** when the system determines that sufficient information has been gathered to support a meaningful response. Retrieved dataset records are provided to the summarization LLM in structured markdown format, along with contextual inputs such as the original query, resolved addresses, GeoScope definitions, and dataset metadata. A carefully designed prompt guides the LLM to identify and emphasize relevant risk signals—such as active versus historical conditions—while avoiding unsupported inference.

Through this process, GeoRisk AI transforms heterogeneous geospatial records into clear, project-manager-friendly summaries grounded explicitly in retrieved evidence. The agent may re-invoke the summarization module in response to follow-up queries (e.g., expanded spatial scope or comparison with another address), allowing summaries to evolve without re-executing the full pipeline unnecessarily.

4.4 Role of the Agentic Framework

By structuring GeoRisk AI as an agentic system, the application achieves several key benefits. First, it enables **adaptive reasoning**, allowing the system to tailor its workflow to each query rather than forcing all interactions through a fixed pipeline. Second, it improves **computational efficiency** by avoiding redundant data retrieval and reusing context across turns. Third, it enhances **user experience and interpretability** by enabling transparent data previews, explaining follow-up suggestions, and maintaining conversational continuity. Thus, altogether, the agentic framework transforms GeoRisk AI from a static query–response tool into an **autonomous geospatial intelligence assistant** capable of multi-step reasoning, dynamic decision-making, and iterative dialogue, thereby laying the foundation for scalable, multi-domain urban risk assessment.

5. System Interface, Capabilities, and Usage

5.1 Application UI, Features, and Deployment

GeoRisk AI is delivered through a modern, web-based chatbot interface designed to make advanced geospatial reasoning accessible to non-expert users. The application features a **dedicated landing page** that introduces system capabilities, provides usage guidance, and serves as an entry point into the conversational experience. From this landing page, users can initiate a chat session through an intuitive, contemporary chatbot UI (see *Appendix A7* for UI screenshots).

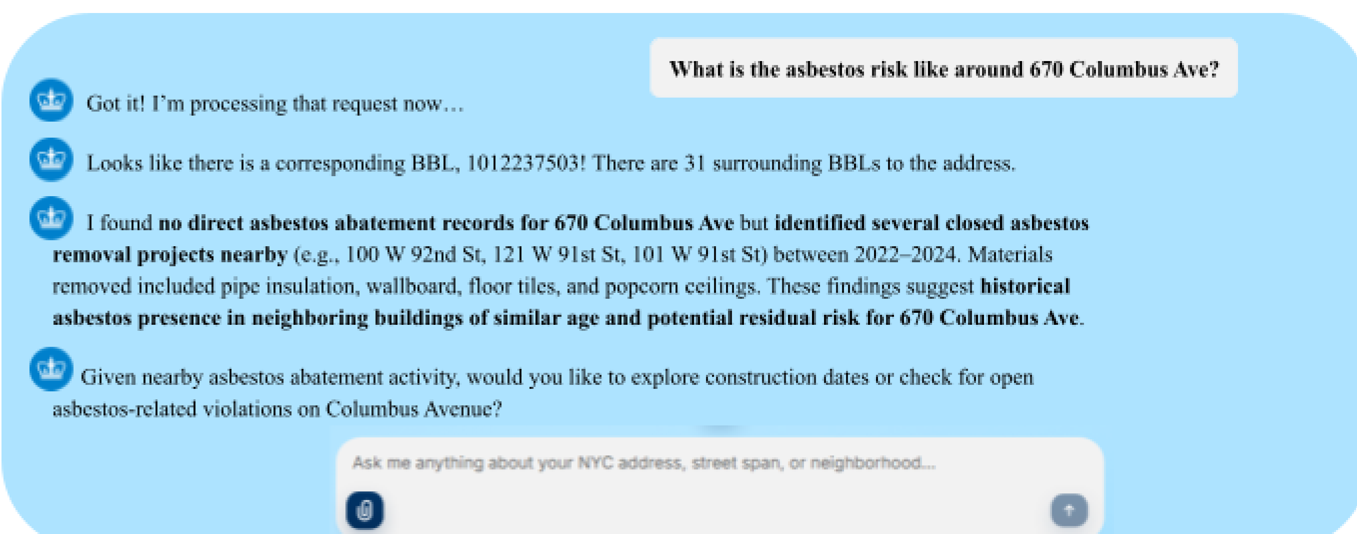
The system supports **user authentication and persistent sessions**, enabling users to create accounts and log in securely. Each user's chat history is stored in a lightweight SQLite-backed database, allowing conversations to persist across sessions. This design enables users to seamlessly resume prior analyses, reference earlier results, and build multi-turn investigations over time without losing context. Additionally, to ensure responsive performance, GeoRisk AI employs a **warm-start initialization strategy**. Upon application startup, large local datasets are preloaded and indexed, and an initial warm-up call is issued to the LLM backend. This significantly reduces latency for subsequent user queries, allowing the agentic workflow to execute efficiently even when operating over large, multi-domain geospatial datasets.

Moreover, the frontend is fully configurable and styled with Columbia University branding, demonstrating how the interface can be adapted for institutional deployment. The application is built using the Python package Chainlit, enabling rapid iteration and flexible deployment. GeoRisk AI can be run locally in a Python environment for development and research purposes, or deployed as a hosted web application. The LLM backend is modular and model-agnostic, allowing different language models to be substituted with minimal changes to the surrounding system.

As a whole, these interface and deployment choices transform a complex agentic geospatial reasoning system into a practical, user-friendly application that supports real-world urban risk analysis workflows.

5.2 Example: Ambiguous Query (Condensed Output)

The example below shows the chatbot's **ability to handle complex, ambiguous user input and synthesize a nuanced, context-aware response.** When asked, *"What is the asbestos risk like around 670 Columbus Ave?"*, the system interprets the vague spatial intent ("around"), locates the nearest geounits through the Geoclient API using only a partial input (house number + street), and identifies relevant nearby asbestos abatement records even when no direct matches exist for the queried address. The generated summary contextualizes these findings by highlighting historical asbestos activity, proximity-based residual risk, and relevant neighboring projects while suggesting meaningful follow-up questions to continue the dialogue.



While this example highlights **advanced reasoning and spatial nuance**, GeoRisk AI also performs efficiently on simpler, direct queries—such as *"Is there asbestos risk at 1633 Broadway?"*—where it can extract address details, locate a precise BBL match, and summarize site-specific information like material type, project status, and timeline with minimal ambiguity. Moreover, in comparative queries such as *"Compare asbestos risk between 405 Lexington Avenue and 1633 Broadway,"* the chatbot can correctly parse multi-address intent and return structured summaries. These capabilities illustrate **the chatbot's flexibility across varying levels of query complexity, from direct address lookups to sophisticated, spatially aware reasoning.** The full LLM-generated output for the shown interaction is provided in *Appendix A1*.

6. Evaluation and Performance Analysis

6.1 Evaluation Framework & Methodology

To validate GeoRisk AI's transition from our mid-term single-domain prototype to a comprehensive urban intelligence agent, we implemented a rigorous, multi-tiered evaluation framework. This approach allowed us to assess the system's ability to orchestrate queries across **20 integrated datasets** spanning six risk categories.

Our testing strategy consisted of two distinct phases:

- **Phase 1 (MVP Baseline):** The midterm evaluation focused on a Ground Truth Set of 20 queries (see *Appendix A5*) specific to the *Asbestos Control Program*. This established a baseline competency in address parsing, data retrieval, and risk summarization.
- **Phase 2 (Final System):** To evaluate the full agentic architecture, we expanded the test suite to **80+ queries** (4 distinct queries per dataset across 20 datasets). These queries stress-tested specific behaviors including direct BBL lookups, complex street span logic (e.g., LION geometry), and multi-turn reasoning with context retention.

6.2 Metric Definition: The GeoRisk Reliability Index (GRI)

We assessed the system's performance using 10 key metrics, aggregated into a weighted grand score—the GeoRisk Reliability Index (0-1), or GRI. This index aggregates the 10 key performance indicators into three weighted pillars: **Pipeline Health**, **Response Quality**, and **Agentic Capability**. The weighting prioritizes pipeline correctness and response accuracy over conversational behavior, reflecting the system's primary objective of delivering reliable, data-grounded risk insights, with agentic interaction serving as a secondary but complementary capability.

The index is calculated as follows:

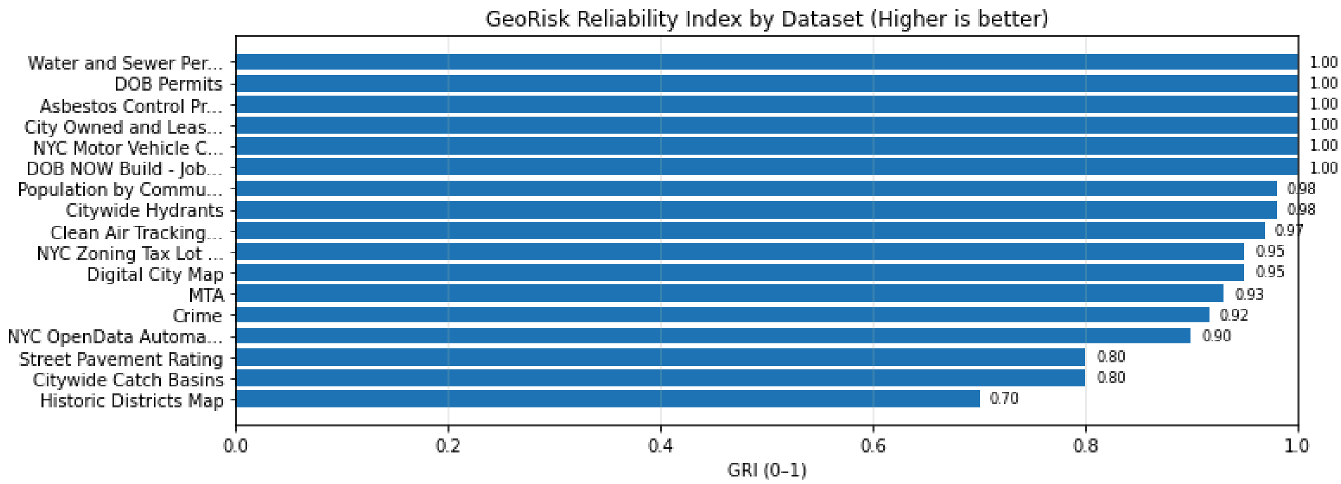
$$GRI = (0.4 \times Pipeline) + (0.4 \times Quality) + (0.2 \times Agentic)$$

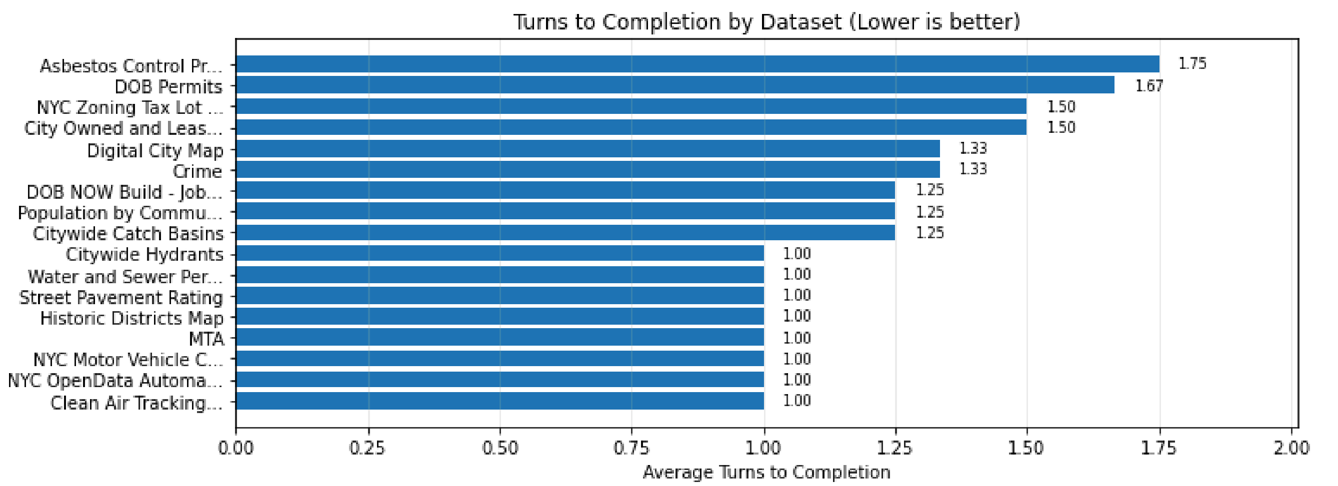
Pillars & Component Metrics:

1. **Pipeline Health** (Weight: 0.4): **Parsed Address Accuracy** and **Parsed Dataset Accuracy** measure the correct identification of user intent. **BBL Accuracy** and **Adapter Accuracy** evaluate geospatial standardization success (e.g., mapping addresses to BBLs or NTA codes). **Filter Validity** ensures the generated filtering logic is syntactically correct.
2. **Response Quality** (Weight: 0.4): **Record Match Accuracy** confirms that retrieved data rows align with user constraints. **Summary Faithfulness** is a pass/fail check preventing hallucinations. **Conversation Quality** rates the response's overall coherence.
3. **Agentic Capability** (Weight: 0.2): **Follow-up Solvability**, and **Follow-up Question Asked** measure the agent's ability to persist context across turns and proactively guide the user for relevant next steps. We analyzed **Turns to Completion** separately from this index to evaluate the computational cost of queries independently of their accuracy.

6.3 Performance Analysis

Comparing the final GeoRisk AI system against the midterm MVP baseline highlights a substantial architectural evolution. While the MVP achieved high accuracy on a single dataset (Asbestos Control Program) using static, single-turn address lookups, the final system preserves this level of reliability while scaling to **20 heterogeneous datasets** and supporting significantly more complex geospatial queries. The following visualizations demonstrate that the transition from a rigid pipeline to an agentic framework did not degrade performance despite increased geospatial and reasoning complexity.





The above figures summarize GeoRisk AI's end-to-end performance across a representative subset of evaluated datasets, capturing both **response quality** (GRI) and **conversational efficiency** (turns-to-completion). As shown in the GRI plot, the agent consistently produces high-quality, evidence-grounded summaries, with modest degradation only for datasets that require longer or more ambiguous queries for broader spatial or multi-dataset reasoning. The Turns to Completion plot shows that most queries are resolved within a single conversational turn, while datasets requiring additional turns—such as permits and zoning data—consistently retain high GRI scores. This indicates that increased conversational depth is used productively to refine the analytical scope or spatial context, rather than to compensate for reduced accuracy. Complete evaluation data and results are provided in *Appendix A8*.

Overall, this evaluation demonstrates that GeoRisk AI's agentic architecture scales reliably from a single-domain prototype to a multi-domain urban intelligence system **without sacrificing accuracy or efficiency**. By maintaining high response quality across heterogeneous datasets while selectively increasing conversational depth only when analytical complexity demands it, the system exhibits robust agentic behavior rather than brittle pipeline execution. These results confirm that GeoRisk AI can autonomously coordinate geospatial reasoning, dataset selection, and multi-turn interaction in a principled and performant manner, establishing a strong foundation for real-world deployment in urban risk assessment and decision support contexts.

6.4 Next Steps

Future work will focus on extending GeoRisk AI's coverage, depth, and practical utility for real-world deployment. A primary direction is the **integration of proprietary and internal datasets provided by project mentors**, including non-public risk indicators and project-specific data. Incorporating these sources would significantly enhance the system's relevance and enable higher-stakes decision support beyond what is possible using publicly available datasets alone.

In parallel, GeoRisk AI can be extended from interactive analysis to **structured risk report generation**. Building on its agentic conversational capabilities, the system could generate comprehensive, project-ready risk reports that synthesize findings across multiple domains, locations, and time horizons. Through conversational refinement, users would be able to tailor report scope, detail, and emphasis (e.g., environmental versus construction risk), allowing the chatbot to function as both an exploratory analysis tool and a report authoring assistant.

Finally, insights from the evaluation in Section 6.3 motivate **further refinement of agentic control**. While the system achieved consistently high GeoRisk Reliability Index (GRI) scores, modest performance variation for broader spatial scopes and multi-dataset reasoning suggests opportunities for improvement. Future work will leverage evaluation signals—such as parser confidence scores and turns-to-completion—to enable proactive clarification, adaptive GeoScope expansion, and dynamic dataset selection. These enhancements further position GeoRisk AI as a scalable foundation for autonomous, end-to-end urban risk intelligence that bridges interactive querying, expert reasoning, and formal decision support.

7. Conclusion

This project demonstrates the feasibility and practical value of an **agentic, location-aware GenAI system** for urban risk assessment in New York City. Through collaboration with Town+Gown: NYC @ DDC, GeoRisk AI advances beyond a single-domain prototype to a fully integrated geospatial intelligence assistant capable of reasoning across more than 20 NYC Open Data sources spanning environmental, zoning, construction, transportation, and public safety risk domains.

At the core of the system is a unified **Borough–Block–Lot (BBL) geospatial backbone**, supported by **MAPPLUTO, LION, and the NYC Geoclient API**, enabling consistent spatial alignment across datasets originally indexed by diverse geounits. Building on this foundation, GeoRisk AI introduces an **agentic architecture** in which large language models act as autonomous decision-makers—interpreting user intent, selecting datasets, determining spatial scope, and orchestrating multi-step geospatial workflows. This design allows the system to adapt flexibly to ambiguous, multi-address, and multi-turn queries without requiring prior geospatial expertise from users.

Evaluation results confirm that the transition from a static, midterm MVP to an agentic, multi-domain system does not sacrifice reliability or accuracy. Across 100+ test queries spanning 20 datasets, GeoRisk AI consistently achieves **high GeoRisk Reliability Index (GRI) scores**, demonstrating s**trong pipeline health, faithful data-grounded summarization, and effective agentic behavior**. Increased conversational depth is shown to reflect productive refinement of spatial context rather than error recovery, underscoring the effectiveness of the agentic control framework.

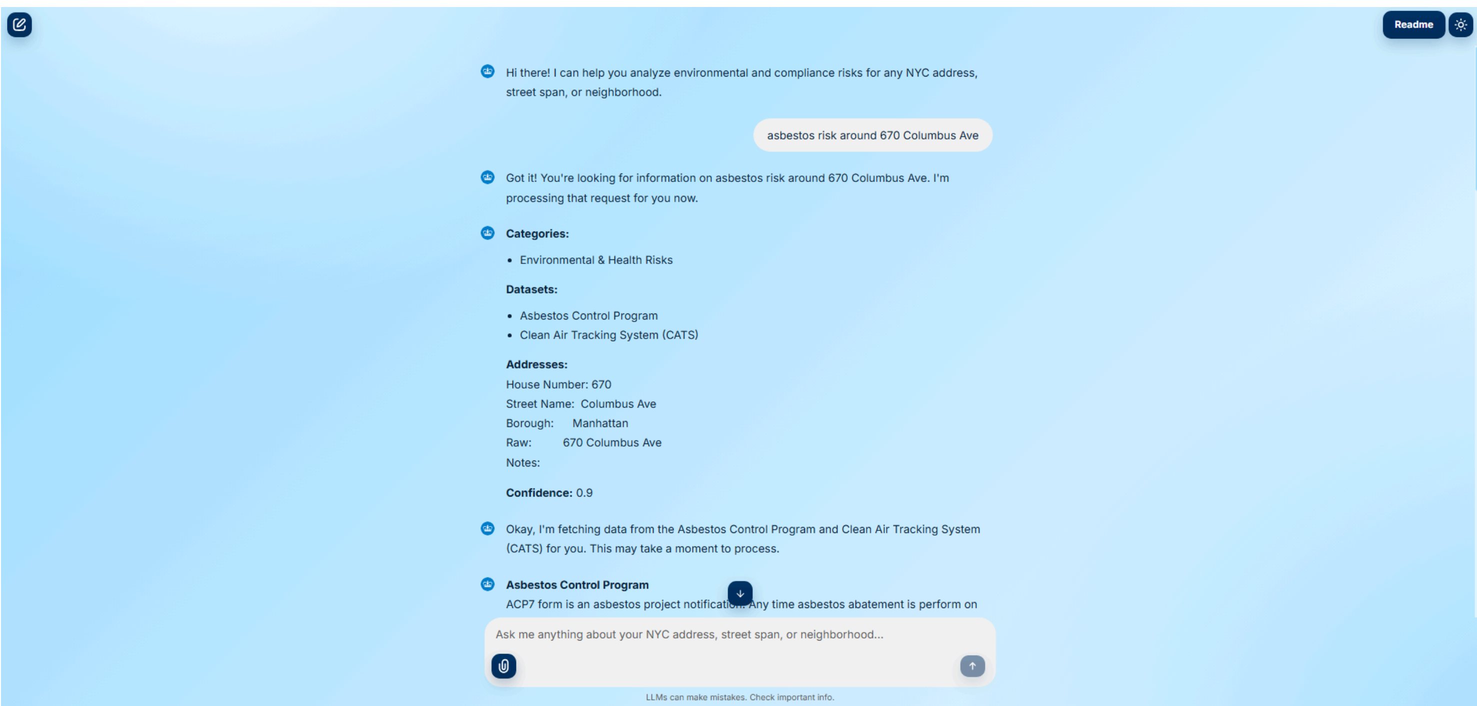
Thus, GeoRisk AI bridges deterministic geospatial computation with generative reasoning to deliver clear, evidence-based, and context-aware urban risk insights. By transforming fragmented public datasets into an interactive conversational decision-support experience, the system establishes a **scalable foundation** for real-world deployment within city agencies. Ultimately, this Capstone project illustrates how agentic GenAI systems can fundamentally improve the accessibility, consistency, and transparency of infrastructure risk assessment across the city's built environment.

8. Contributions

Each team member played a key role in designing and implementing the MVP. All five members collaboratively performed exploratory data analysis across more than 20 NYC Open Data sources and co-developed the bidirectional geounit adapters for geospatial standardization. Louis led the development of the LLM Query Parser and the DataHandler module, overseeing dataset retrieval and integration. Kevin designed and implemented the GeoScope mechanism that automates spatial context generation and buffering. Max developed the Risk Summarization pipeline, creating structured prompts and conducting extensive prompt engineering for the LLM-based analysis. Sharon built the evaluation framework, leading all testing and evaluation efforts to assess overall application performance and quality. Rishi designed the overall agentic system architecture, orchestrating the multi-stage pipeline, UI design, and integration of the conversational interface that connects all components into a cohesive end-to-end chatbot system.

9. Appendix

A1: Full Ambiguous Query Chatbot Conversation



ACP7 form is an asbestos project notification. Any time asbestos abatement is perform on quantities greater than a minor project amount, the applicant has to file this form with DEP Asbestos Control Program (ACP). All asbestos documents are filed through the Asbestos Reporting and Tracking System (ARTS) E-file system. This system is web based and entirely paperless. All information on the ACP7 is essential to meet the requirements setforth in the asbestos rules and regulations Title15, Chapter 1 (RCNY). ACP enforcement staff utilizes this form for inspection of asbestos abatement activities.

Preview:

tru	start_date	end_date	status_description	street_activity
TRU1849MN22	2022-07-15T00:00:00.000	2023-07-21T00:00:00.000	Closed	No
TRU0120MN23	2023-03-20T00:00:00.000	2024-01-19T00:00:00.000	Closed	No
TRU1374MN23	2023-10-05T00:00:00.000	2024-06-07T00:00:00.000	Closed	No
TRU1374MN23	2023-10-05T00:00:00.000	2024-06-07T00:00:00.000	Closed	No
TRU2897MN23	2024-01-24T00:00:00.000	2024-11-19T00:00:00.000	Closed	No

Ask me anything about your NYC address, street span, or neighborhood...

LLMs can make mistakes. Check important info.

| TRU2897MN23 | 2024-01-24T00:00:00.000 | 2024-11-19T00:00:00.000 | Closed | No |

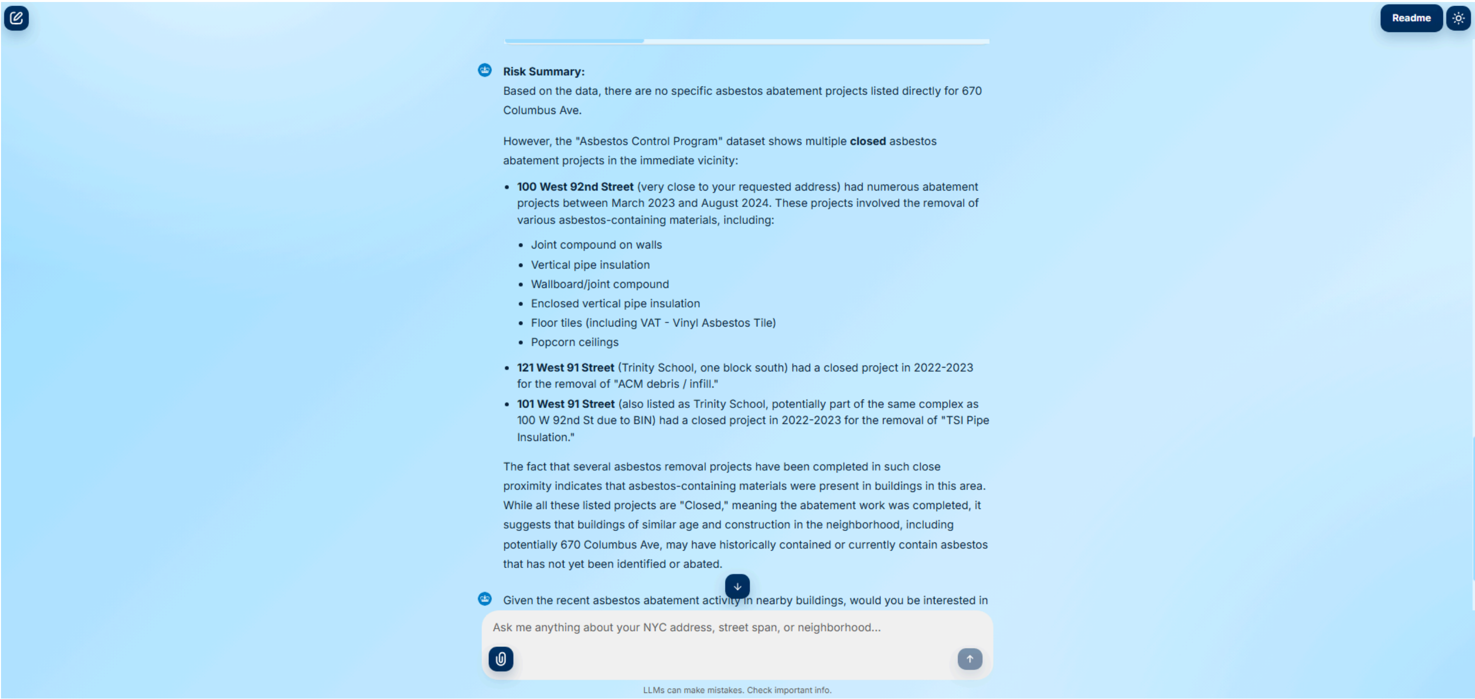
Clean Air Tracking System (CATS)

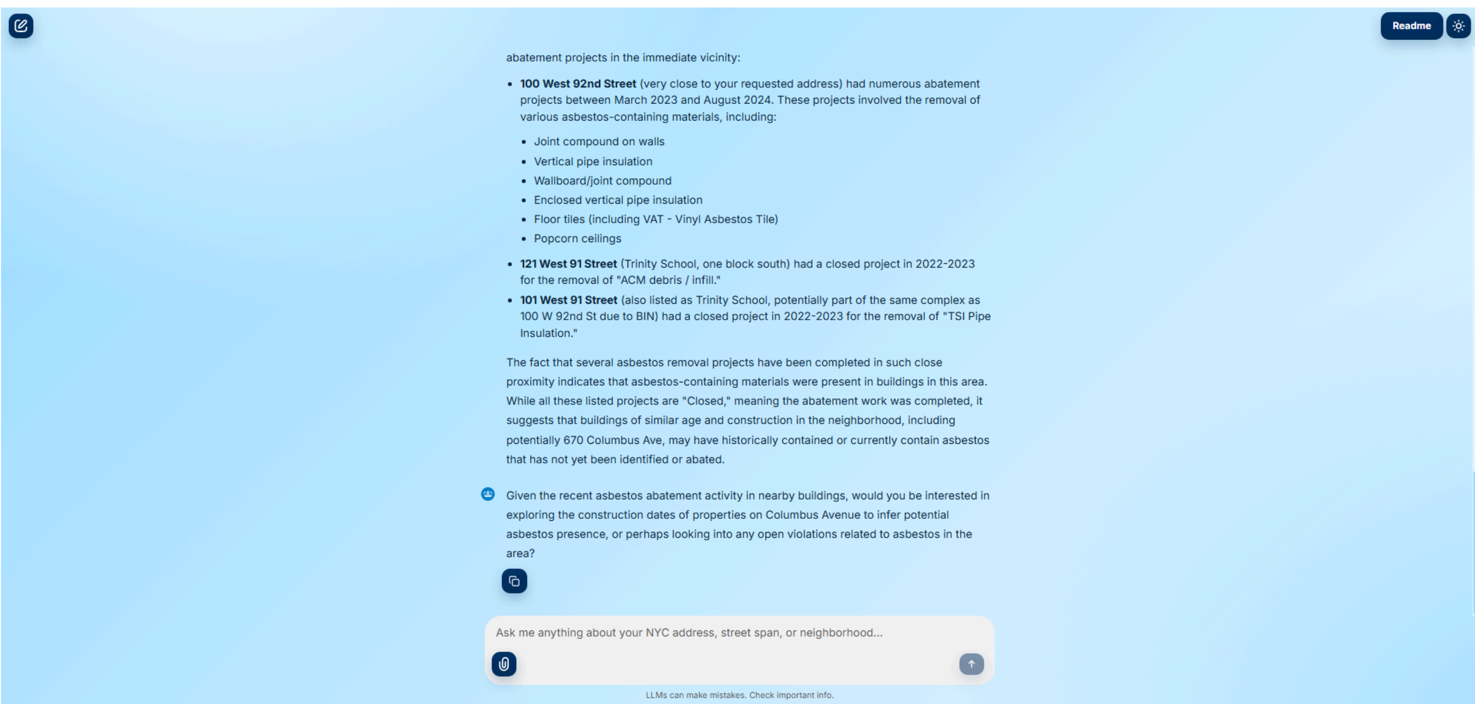
Preview:

requestid	applicationid	requesttype	house	street	borough	
54696	CA059089	REGISTRATION	142	FULTON STREET	MANHATTAN	1001
378827	CA022047	CERTIFICATE TO OPERATE	70	nan	MANHATTAN	3000
394994	CR081417	REGISTRATION	1581	FULTON AVENUE	BRONX	2009
118102	CA026270	CERTIFICATE TO OPERATE	2419	DAVIDSON AVENUE	BRONX	2014
456474	CA062298	REGISTRATION	19-47	80 STREET	QUEENS	4444

Ask me anything about your NYC address, street span, or neighborhood...

LLMs can make mistakes. Check important info.





A2: Data Descriptions

Dataset Name	Spatial Unit	Risk Category	Description

NYC OpenData Zoning and Tax Lot Database	BBL	Zoning & Land Use	Contains zoning and land use classifications for each tax lot, used to determine regulatory and development constraints.
NYC OpenData PLUTO	BBL	Zoning & Land Use, Public Safety & Social Context	Provides detailed building and land use data, including lot area, ownership, and construction year.
NYC OpenData Automated Traffic Volume Counts	Lat/Lon, Street	Public Safety & Social Context, Transportation & Traffic	Records traffic volume at bridges and major roads, supporting transportation and congestion analysis.
NYC OpenData Motor Vehicle Collisions	Lat/Lon, Street	Transportation & Traffic, Public Safety & Social Context	Documents traffic incidents and crash locations, used to identify transportation risk areas.
DOB Permits	BBL	Construction & Permitting	Permits for construction and demolition activities in the City of New York. Each record represents the life cycle of one permit for one work type.
Asbestos Control Program	BBL	Environmental & Health Risks	Any time asbestos abatement is performed on quantities greater than a minor project amount, the applicant has to file this form with DEP Asbestos Control Program (ACP).
Digital City Map Shapefile	Lat/Lon, EPSG:4326	Zoning & Land Use	The Digital City Map (DCM) data represents street lines and other features shown on the City Map, which is the official street map of the City of New York.
Historic Districts Map	Polygon	Zoning & Land Use	Defines boundaries of designated historic districts.
LION	EPSG:2263	Zoning & Land Use	A single line street base map representing the city's streets and other linear geographic features.
Zoning GIS Data	EPSG:2263	Zoning & Land Use	This data set consists of 6 classes of zoning features: zoning districts, special purpose districts, special purpose district subdistricts, limited height districts, commercial overlay districts, and zoning map amendments.

Population by Community Districts	N/A	Public Safety & Social Context	Aggregated population by community districts for the years 1970, 1980, 1990, 2000 and 2010.
Population by Neighborhood Tabulation Area	NTA	Public Safety & Social Context, Environmental & Health Risks	Aggregations of census tracts that are subsets of New York City's 55 Public Use Microdata Areas (PUMAs) of change in population for 2000 and 2010
Crime	Precinct	Public Safety & Social Context, Transportation & Traffic	Statistical breakdown by citywide, borough, and precinct.
Street Construction Permits	Lat/Lon, Street	Construction & Permitting, Transportation & Traffic	Over 150 different types of sidewalk and roadway construction permits to utilities, contractors, government agencies and homeowners.
MTA Subway and Train Lines	Lat/Lon	Transportation & Traffic Comparative Site Queries	A dataset listing all NYC subway and Staten Island Railway stations.
City Owned and Leased Property	BBL	Zoning & Land Use, Construction & Permitting	Identifies city-owned and leased properties and their usage.
Sewer System Data	Point / Polygon	Environmental & Health Risks Comparative Site Queries	This MS4 Map represents the known MS4 outfalls and drainage areas as of August 1, 2020 and provides additional data relevant to the MS4 Permit.
Clean Air Tracking System (CATS)	BBL	Environmental & Health Risks	An online platform managed by the NYC Department of Environmental Protection that facilitates end-to-end processing of air quality.

Water and Sewer Permits	BBL	Construction & Permitting Comparative Site Queries	Contain information about the different types of applications approved and permits issued on a regular basis.
DOB NOW: Build – Job Application Findings	BBL	Construction & Permitting Comparative Site Queries	Records building job filings and applications, with the exception of electrical, elevator, and LAA
Citywide Catch Basins	Lon/Lat	Environmental & Health Risks	The Citywide Catch Basins dataset maps over 150,000 catch basins that collect stormwater and direct it into NYC's sewer network or nearby waterbodies, supporting drainage and flood management.
Parks Monuments	Borough	Zoning & Land Use	This data is a table of monuments that is maintained by NYC Parks.
Citywide Hydrants	Lat/Lon	Public Safety & Social Context Comparative Site Queries	Lists all fire hydrant locations across NYC.
Street Pavement Rating	Borough	Transportation & Traffic	The Agency performs ongoing assessment of New York City streets. Ratings are based on a scale from 1 to 10

A3: Geounits

Geounit Type	Description	Example Source Dataset
BBL	Parcel-level property identifier	PLUTO
BIN (Building Identification Number)	Building-level identifier	Geoclient
Latitude/Longitude (EPSG:4326)	Coordinate-based	MTA Subway

Precinct	NYPD / enforcement area	Crime Data
NTA (Neighborhood Tabulation Area)	Demographic aggregation	Population by Neighborhood Tabulation Area

A4: Risk Category

Environmental & Health Risks – asbestos filings, flood zones, noise, and air quality complaints	**Zoning & Land Use** – zoning districts, landmarks, historic designations, land use restrictions	**Construction & Permitting** – active DOB building permits, street construction, asbestos filings
Transportation & Traffic – vehicle collisions, traffic volumes, pedestrian injury patterns	**Public Safety & Social Context** – NYPD complaint types, crime density, demographic context	**Comparative Site Queries** – multi-location comparisons across any of the above categories

A5: Midterm Report Evaluation Set

Bucket 1: Exact Matches
Purpose: To test the baseline functionality (Parser, Adapter, Retriever) on clean, unambiguous queries with known data.

1. Is there asbestos risk at 1633 Broadway, New York, NY 10019?
2. What is the asbestos risk at 405 Lexington Avenue?
3. Tell me about asbestos at 111 8th Avenue, New York, NY.
4. asbestos records for 1 World Trade Center
5. Show me the risk profile for 550 Madison Avenue.

Bucket 2: Spatial Proximity & Imprecise Queries
Purpose: To test the "Geoscope" and "Record Match Accuracy" on ambiguous spatial language.

6. Is there asbestos risk around 670 Columbus Ave, New York, NY 10025?
7. Show me asbestos reports near 80 Centre Street.
8. What is the risk on the same block as 500 West 120th Street?

Bucket 3: Typos & Dirty Data
Purpose: To test the Parsed Address Accuracy and the robustness of the prompt interpreter.

9. asbestos risk for 1633 Broadwy
10. 405 lexingtn ave
11. 111 8th av
12. 23-01 45th road, queens

Bucket 4: Zero-Result & Known Misses
Purpose: To test "Summary Faithfulness" and "Conversational Quality" when no data is found.

13. asbestos at 1 Sheep Meadow, Central Park
14. What is the asbestos risk at 1000 5th Avenue?
15. asbestos records for 123 Main Street, New York, NY

Bucket 5: Complex & Comparative Queries
Purpose: To test the identified MVP limitations and the system's ability to handle multi-intent queries.

16. Compare asbestos risk between 405 Lexington Avenue and 1633 Broadway.
17. Which is older, 1633 Broadway or the Chrysler Building?
18. Show me commercial buildings in Midtown with 'A' energy grades.

Bucket 6: Guardrail & Out-of-Domain Tests
Purpose: To validate the Data Scoping and ensure the bot rejects out-of-domain questions.

19. What is the weather in Morningside Heights today?
20. Who is the mayor of New York City?

A6: Project Poster

GeoRisk AI

An Agentic Geospatial AI Chatbot for Autonomous Multi-Domain Urban Risk Analysis

Rishi Kumar, Louis Chu, Sharon Tsai, Max Sebrechts, Kevin Ke
Terri Matthews, Subhash Tuladhar, Gwendolyn Gillingham, Sarthak Arora

Data Science Institute
COLUMBIA UNIVERSITY

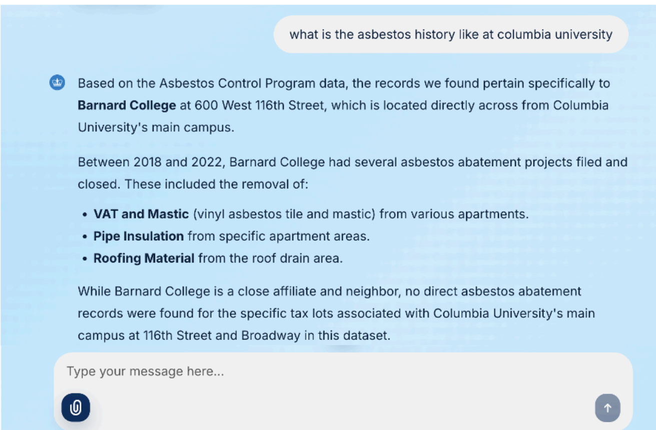
Data Science Capstone Project
Town+Gown: NYC @ DDC Project Controls

Problem Statement

NYC's urban risk data is scattered across dozens of datasets—and nowhere in one place. Project managers must manually reconcile information stored in various geospatial units leading to slow project reviews, inconsistent risk assessments, and additional operational friction.

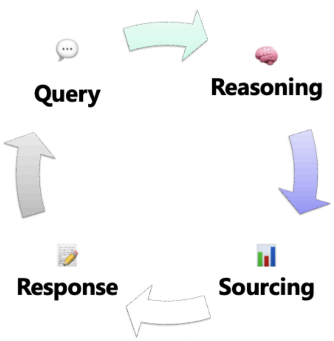
Our Solution

***GeoRisk AI*, autonomous geospatial AI agent that turns natural language queries into relevant, meaningful risk summaries:**

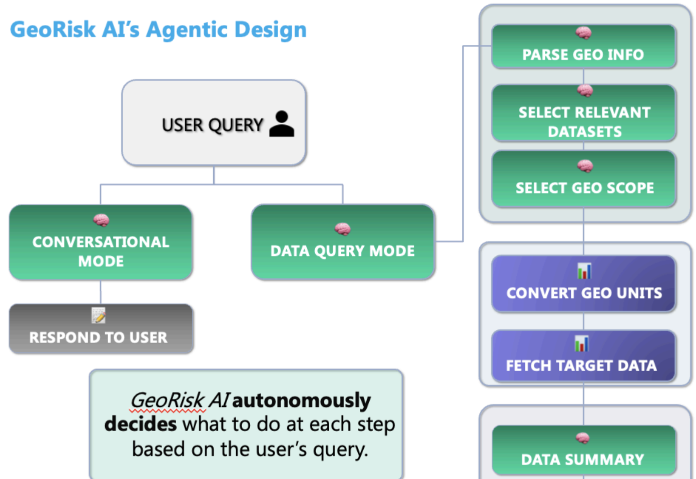


GeoRisk AI's Features

✅ Autonomous Query Parsing	✅ Dynamic Dataset Selection
✅ Contextual Memory	✅ Clear Risk Insights
✅ Adaptive Spatial Awareness	✅ Self-Directed Data Retrieval
✅ Human-Like Follow-Ups	✅ Modern Chatbot UI



GeoRisk AI's Agentic Design



Evaluation Framework

End-to-End Pipeline Performance

Spatial Accuracy	Data Retrieval
Response Quality	Turns to Resolution

Takeaways

By shifting from manual lookup to agentic decision-making, *GeoRisk AI* shows that autonomous systems can independently interpret human queries, select and join relevant data, and generate actionable urban risk insights with minimal human effort.

Acknowledgments

Generative AI tools (ChatGPT) were used for phrasing and clarity in accordance with course policy; all analysis and conclusions are our own. We thank Town+Gown: NYC @ DDC for their guidance and support.

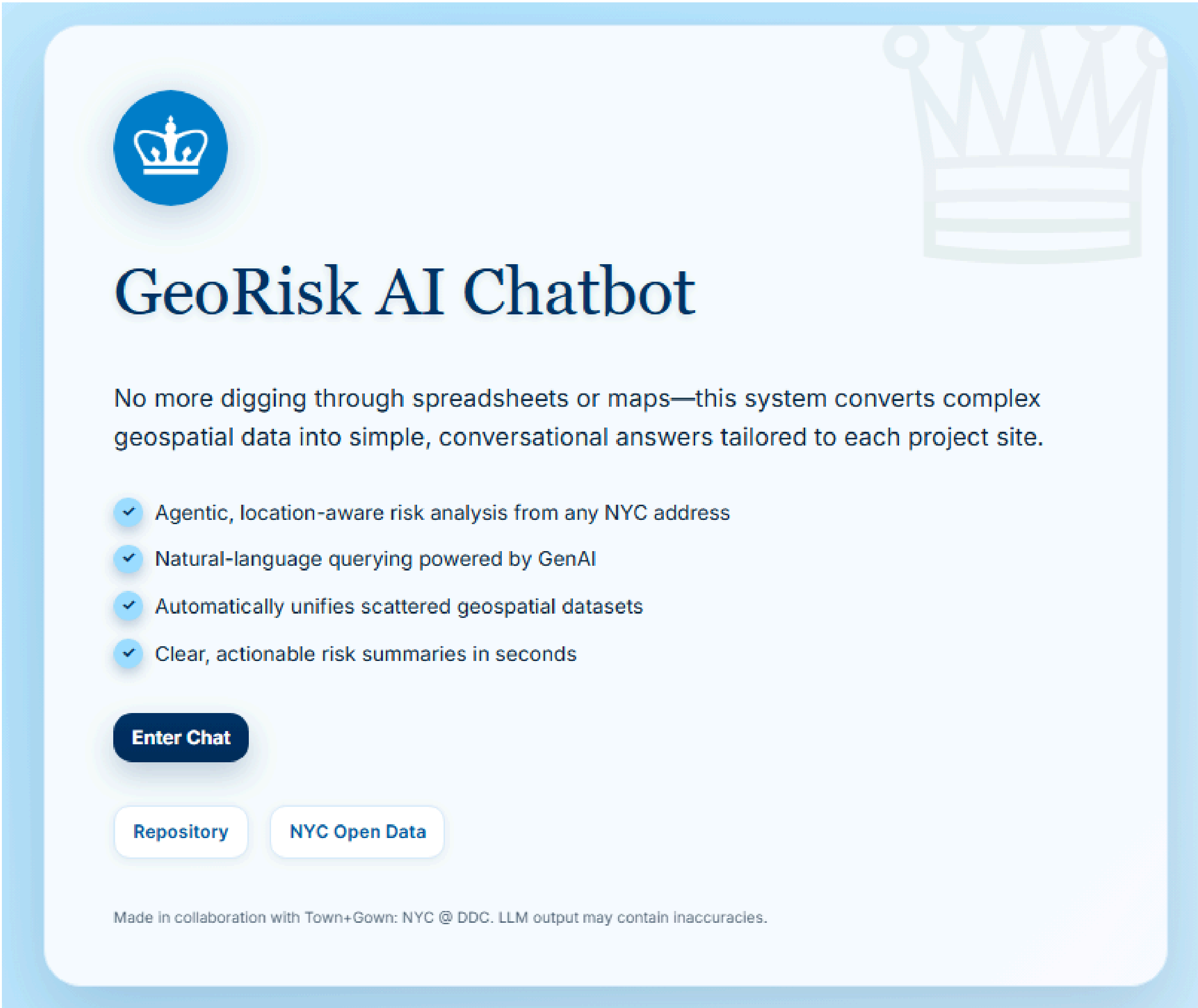
References

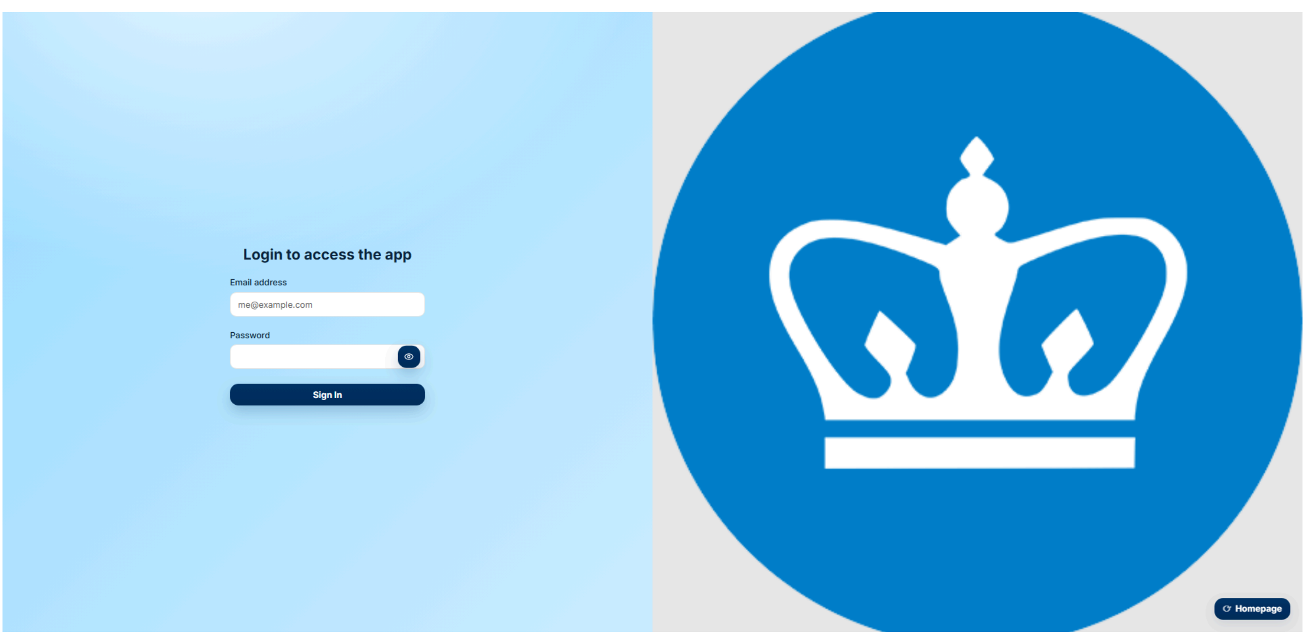
[1] X. Zhu, R. Y. M. Li, M. J. C. Crabbe, and K. Sukpascharoen, "Can a chatbot enhance hazard awareness in the construction industry?," Frontiers in Public Health, vol. 10, Nov. 2022. doi: 10.3389/fpubh.2022.993700

[2] A. Polo-Rodríguez, L. Fiorini, E. Rovini, F. Cavallo, and J. Medina-Quero, "Enhancing Smart Environments with Context-Aware Chatbots using Large Language Models," arXiv preprint arXiv:2502.14469, Feb. 2025. [Online]. Available: https://arxiv.org/abs/2502.14469

[3] K. Tian, E. Mitchell, A. Zhou, A. Sharma, R. Rafailov, H. Yao, C. Finn, and C. D. Manning, "Just Ask for Calibration: Strategies for Eliciting Calibrated Confidence Scores from Language Models Fine-Tuned with Human Feedback," arXiv preprint arXiv:2305.14975, May 2023. Available: https://arxiv.org/abs/2305.14975

A7: UI Screenshots





A8: Evaluation Data

dataset	avg_score	avg_turns
Clean Air Tracking System	0.96875	1
Asbestos Control Program	1	1.75
City Owned and Leased Property	1	1.5
Citywide Catch Basins	0.8	1.25
Citywide Hydrants	0.98	1
Crime	0.9166666667	1.333333333
DOB NOW Build - Job Application	1	1.25
DOB Permits	1	1.666666667
Digital City Map	0.95	1.333333333
Historic Districts Map	0.7	1
MTA	0.93	1
NYC Motor Vehicle Collisions	1	1
NYC OpenData Automated Traffic	0.9	1

NYC Zoning Tax Lot Database	0.95	1.5
Population by Community District	0.98	1.25
Street Pavement Rating	0.8	1
Water and Sewer Permits	1	1

10. Acknowledgements

In accordance with course policy, Generative AI tools, including ChatGPT and Cursor, were used to support the writing of this report. These tools helped refine phrasing, enhance clarity, and condense content. All analyses, interpretations, and conclusions were independently developed and validated by our Capstone student team.

We also want to express our gratitude to our mentors from **Town+Gown: NYC @ DDC** for their invaluable guidance and feedback throughout this project.

11. References

[1] X. Zhu, R. Y. M. Li, M. J. C. Crabbe, and K. Sukpascharoen, "Can a chatbot enhance hazard awareness in the construction industry?," *Frontiers in Public Health*, vol. 10, Nov. 2022. doi: 10.3389/fpubh.2022.993700

[2] A. Polo-Rodríguez, L. Fiorini, E. Rovini, F. Cavallo, and J. Medina-Quero, "Enhancing Smart Environments with Context-Aware Chatbots using Large Language Models," *arXiv preprint* arXiv:2502.14469, Feb. 2025. [Online]. Available: https://arxiv.org/abs/2502.14469

[3] K. Tian, E. Mitchell, A. Zhou, A. Sharma, R. Rafailov, H. Yao, C. Finn, and C. D. Manning, "Just Ask for Calibration: Strategies for Eliciting Calibrated Confidence Scores from Language Models Fine-Tuned with Human Feedback," *arXiv preprint arXiv:2305.14975*, May 2023. Available: https://arxiv.org/abs/2305.14975

[4] R. Brachman, E. K. Morris, A. C. Head, M. A. Hearst, and M. S. Bernstein, "Understanding User Mental Models of Agentic AI Chatbots," *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*, 2025. doi: 10.1145/3708359.3712071