UNIVERSITY of HULL

StructureGPT

Multi-Model RAG System for UK Building Regulations

MSc AI & Data Science University of Hull

Samuel Datubo Jaja

Supervisor: Dr. Aarzoo Dhiman

May 2025

MSC. RESEARCH PROJECT, CENTRE OF EXCELLENCE FOR DATA SCIENCE, ARTIFICIAL INTELLIGENCE & MODELLING (DAIM) UNIVERSITY OF HULL, UK

Agenda

- Introduction
- About the Data
- Methodology
- Live Demo
- Results
- Discussion
- Limitations
- Future Work
- Summary

INTRODUCTION

UK Building Regulations are legally enforced technical standards that govern the design, construction, and alteration of buildings.

Current Information Retrieval Challenges

UK Building Regulations are complex, updated frequently, and exist in static PDFs

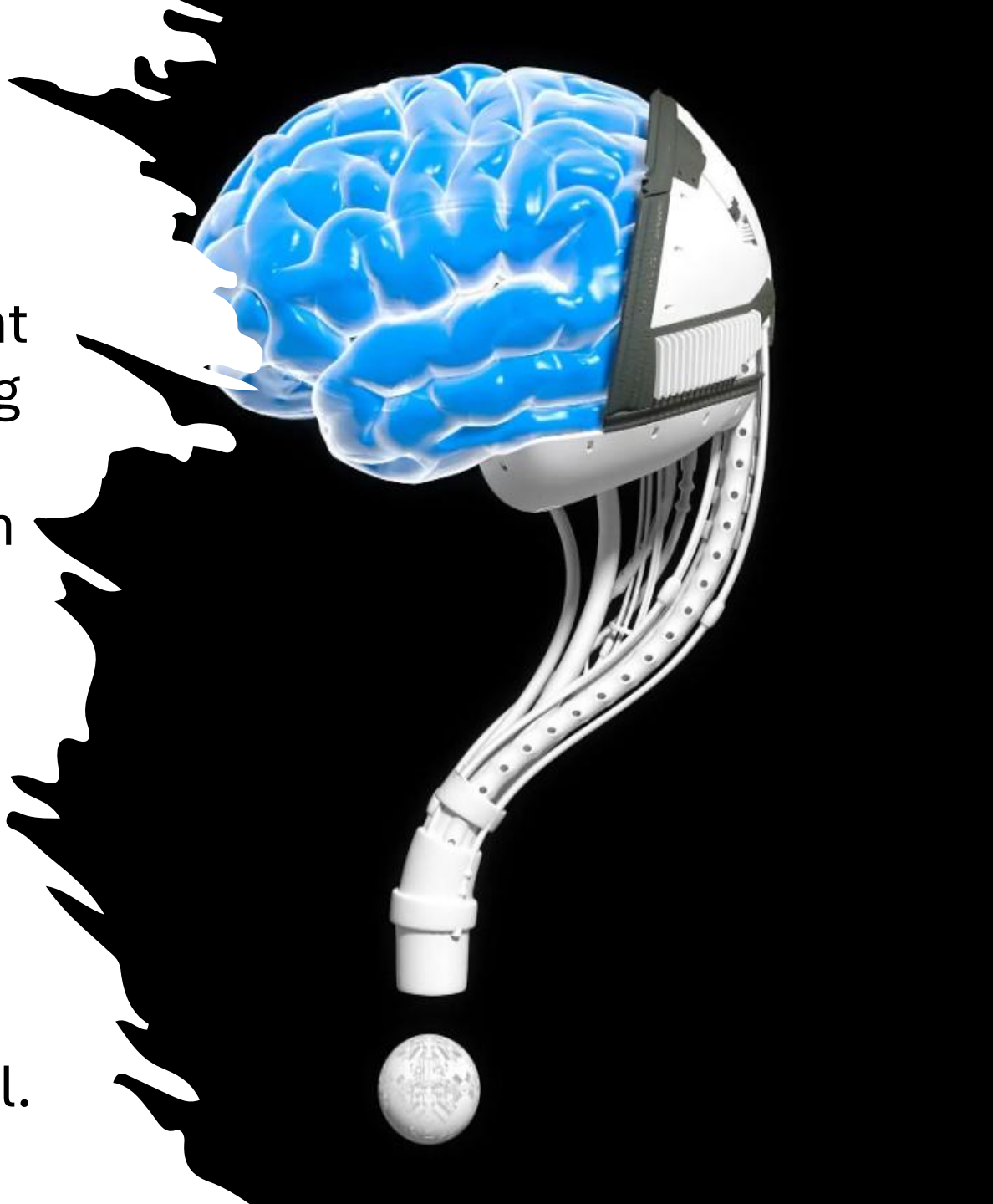
Inspectors, Contractors, surveyors, and planners face difficulties in accessing regulatory answers.

Need for an AI solution that makes these regulations searchable, understandable, and accurate.

RESEARCH QUESTION AND MOTIVATION

- This research investigates how different LLM architectures perform in retrieving UK Building Regulations, focusing on trade-offs between model size, domain specialization through fine-tuning, and deployment architecture.

- It is motivated by limited AI use in navigating complex and frequently updated regulations, and the need for faster, safer access to critical building information through intelligent retrieval.



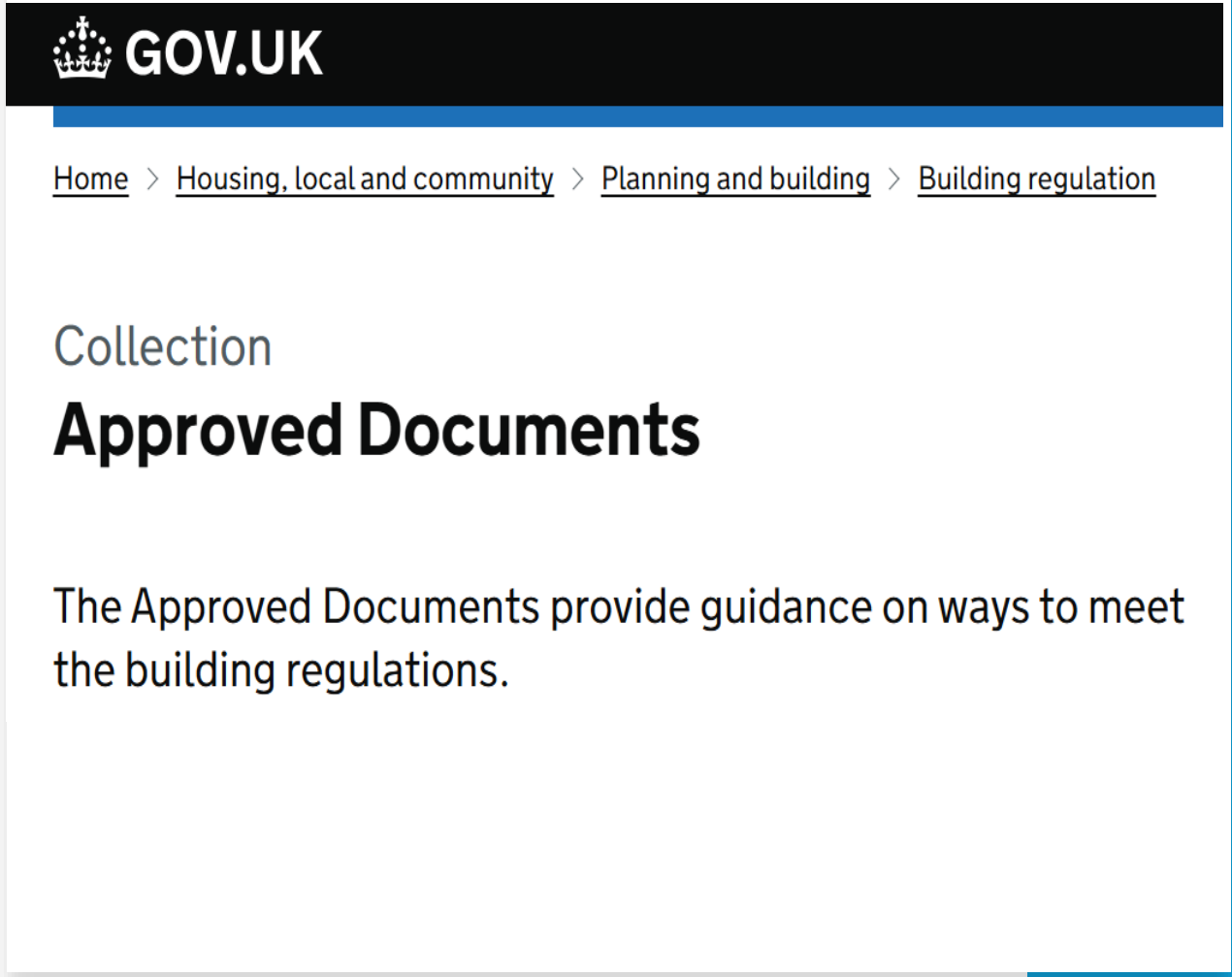
RESEARCH AIM & OBJECTIVES

The primary aim of this research is to enhance a chatbot that provides guidance on UK Building Regulations through the integration of RAG and LLMs. Specific objectives include:

- Expanding the chatbot's knowledge base using PDFs from GOV.UK.
- Implementing RAG to improve response accuracy in chatbot interactions.
- Fine-tuning LLMs with domain-specific data to enhance their contextual understanding and response generation.

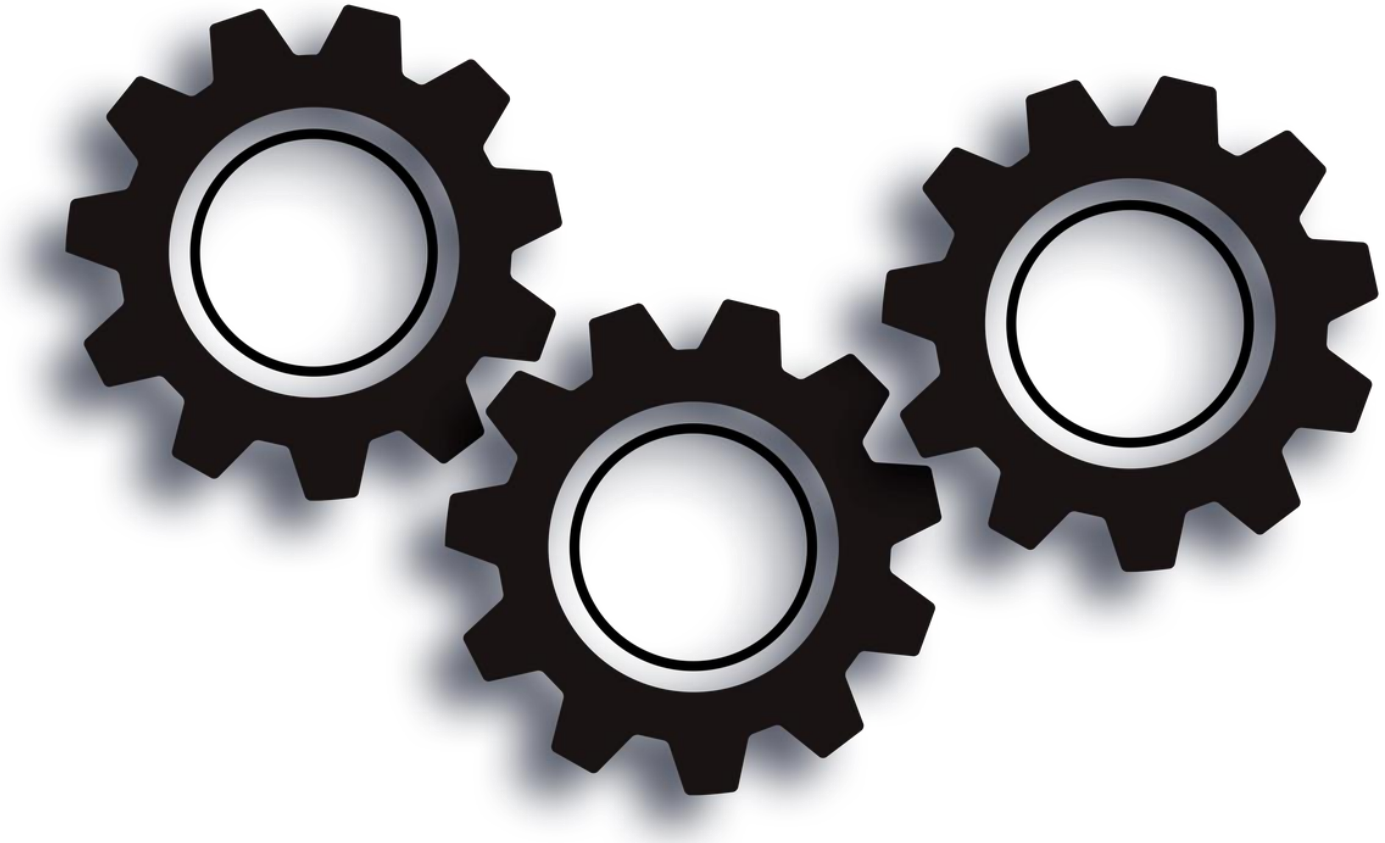
About the Data (Unstructured for RAG & Structured for LoRA Tuning)

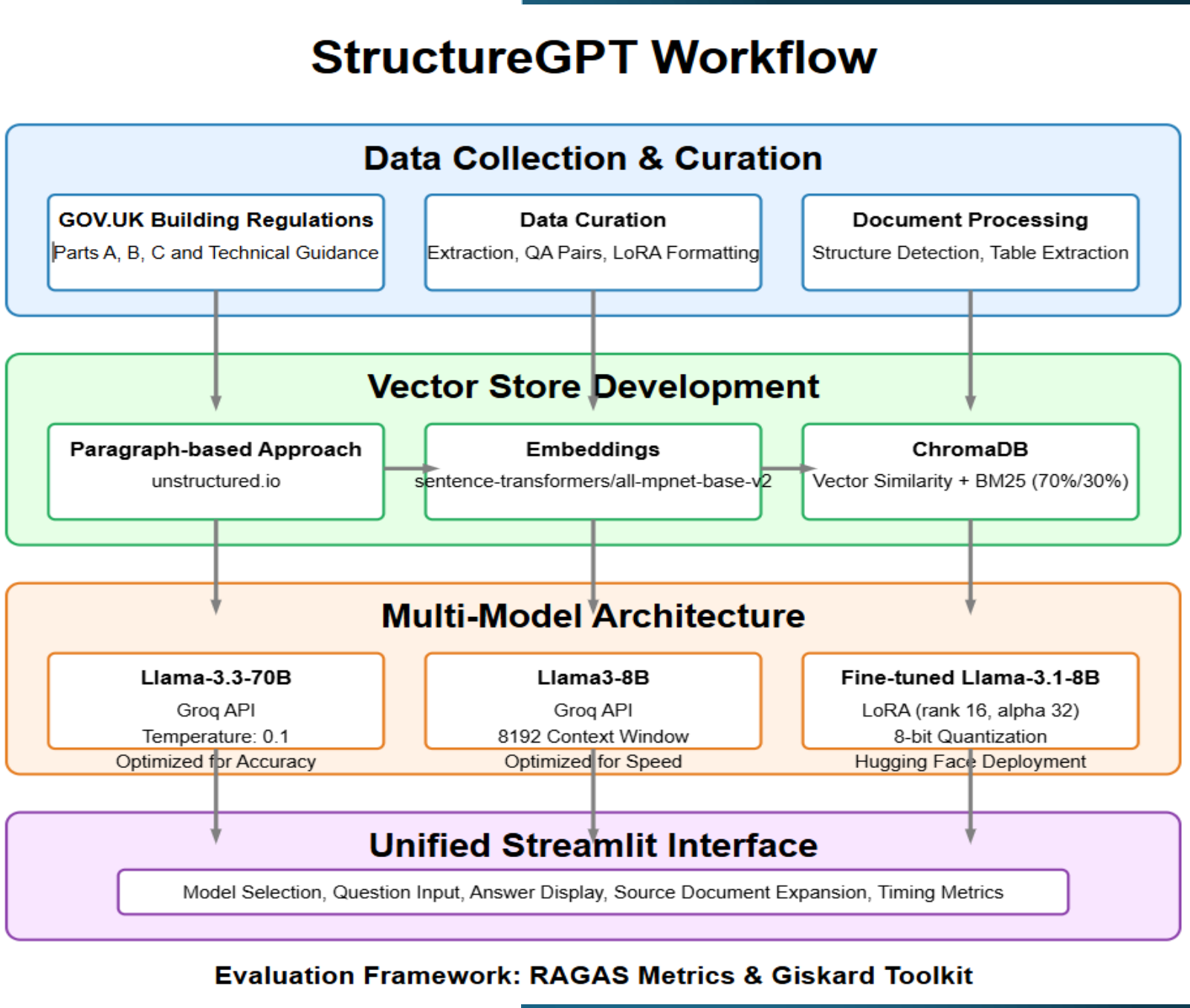
- ✓ **Preprocessing**
- ✓ **18 PDFs [Parts A-S]**
- ✓ **10 FAQ PDFs**
- ✓ **Curated 3000 Instruction-Pair QA**



Source: https://www.gov.uk/government/collections/approved-documents

METHODOLOGY





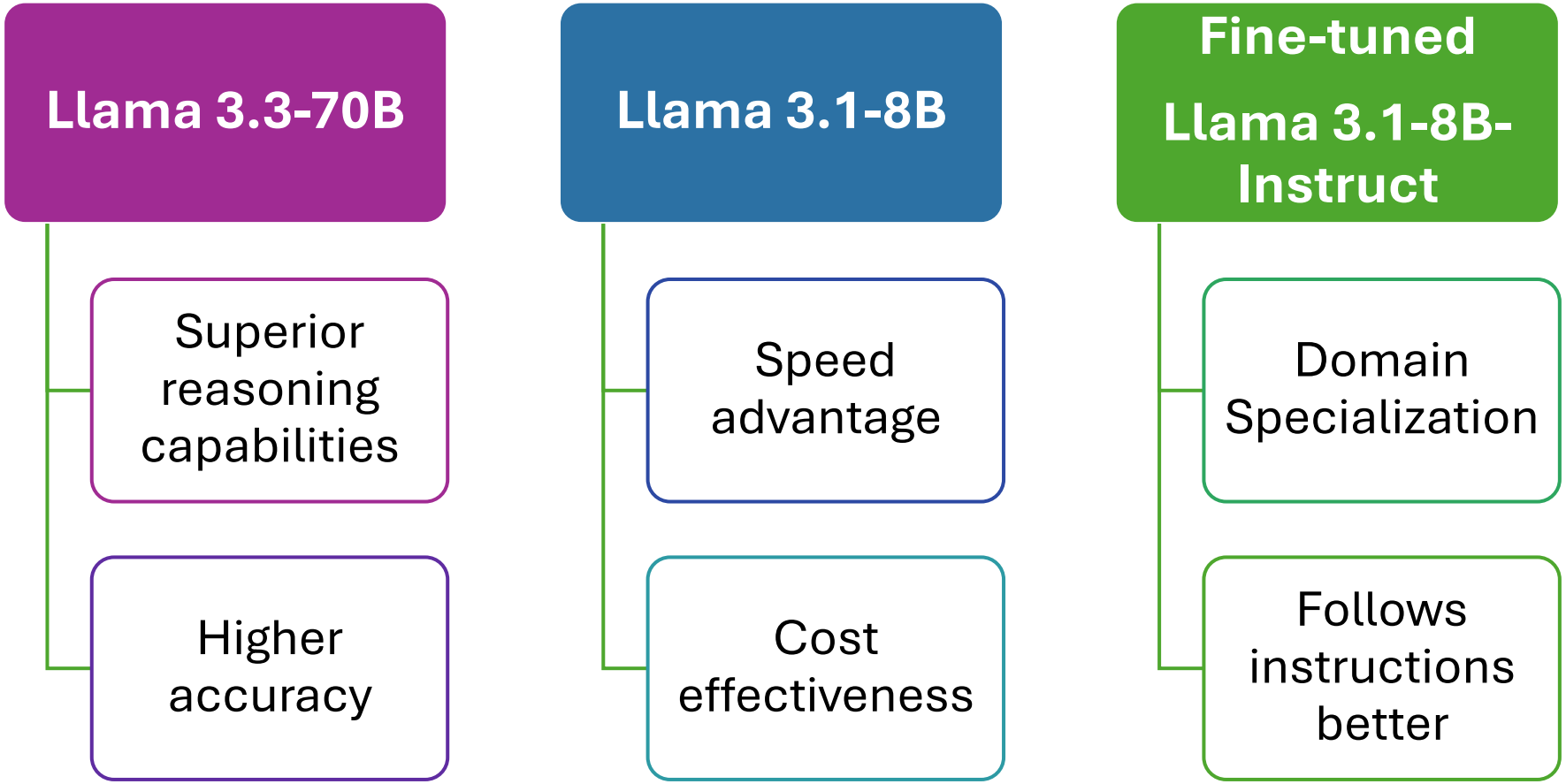
Tools

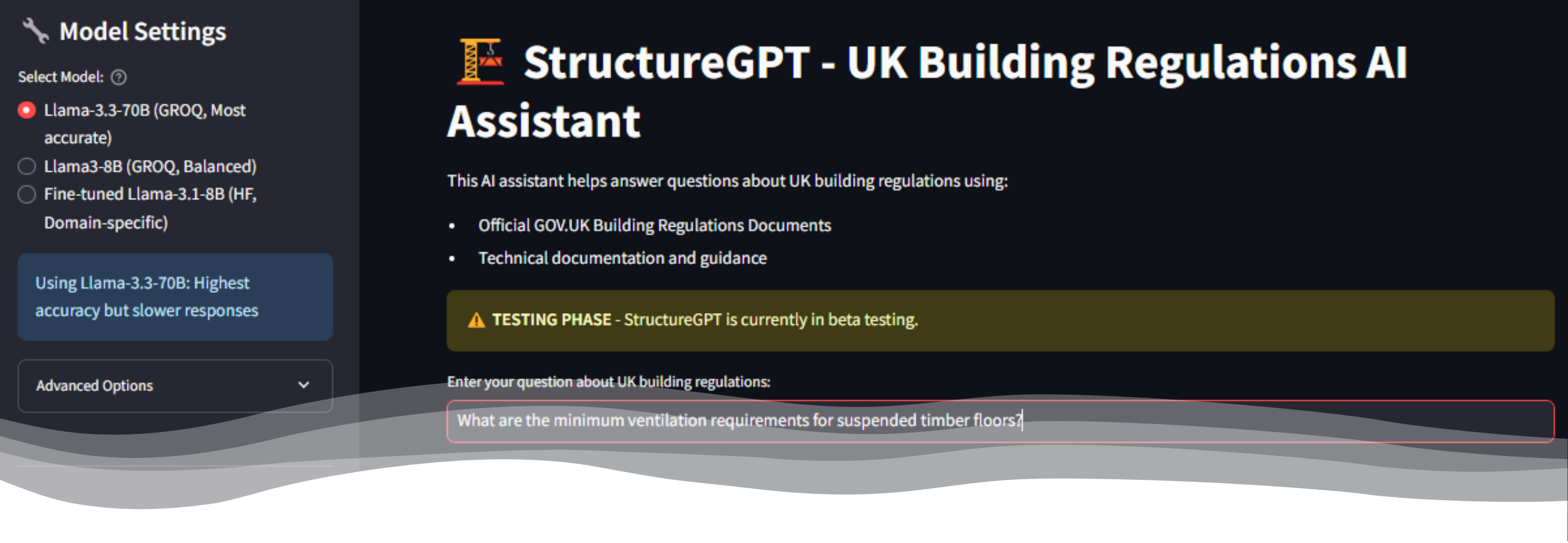
Unstructured API
ChromaDB
Groq API
Streamlit
Hugging Face

Key Terms:
RAG, Low Rank Adaptation & Quantization

Retrieval-Augmented Generation (RAG)	Enhances LLM outputs by retrieving relevant text (e.g., UK Building Regulations) before generating a response, improving accuracy and grounding.
Low-Rank Adaptation (LoRA)	A lightweight fine-tuning method that adds small trainable layers to large models, making them domain-specific with less compute and storage.
Quantization	Reduces model size and speeds up inference by lowering weight precision (e.g., 32-bit → 8-bit), enabling deployment on resource-limited systems.

Large Language Models utilized in RAG System





🔴 LIVE DEMO

https://huggingface.co/spaces/SamuelJaja/uk-building-regulations-bot


RESEARCH
RESULTS
C

Table 1: RAGAS Metrics Comparison

Model	Context Recall	Faithfulness	Factual Correctness
Llama-3.3-70B (GROQ)	0.8500	0.5516	0.5630
Llama3-8B (GROQ)	0.8500	0.4732	0.6880
Fine-tuned Llama-3.1-8B	0.8500	0.2803	0.1420

RAGAS Evaluation

Table 2: Giskard RAG Evaluation Results

Component	Fine-tuned Llama-3.1-8B	Llama3-8B (GROQ)	Llama-3.3-70B (GROQ)
GENERATOR	20.0%	80.0%	60.0%
RETRIEVER	50.0%	100.0%	50.0%
REWRITER	0.0%	100.0%	50.0%
ROUTING	100.0%	100.0%	100.0%
KNOWLEDGE BASE	100.0%	100.0%	100.0%
OVERALL CORRECTNESS	20.0%	80.0%	60.0%

These results show several unexpected patterns:

Giskard Evaluation

Discussion

- ✓ Hybrid retrieval’s consistency across models confirms its strength and potential in other regulatory fields like healthcare or finance.

- ✓ The fine-tuned model’s results indicate that standard LoRA may be insufficient for technical domains. Multi-stage tuning may be needed for formatting, citations, and specification handling.

- ✓ 8B model's performance suggests that smaller models on optimized infrastructure can outperform larger or fine-tuned ones for regulatory tasks.

Limitations

- ✓No multimodal input (diagrams) based on user study
- ✓Limited fine-tuning dataset (~3K QA Pairs)
- ✓Domain Knowledge in Building Construction
- ✓Time Factor

Future Works

Implications, Commercialization & Future Work

- ✓ Add multimodal capabilities (images/diagrams)
- ✓ Better Multi-stage fine-tuning with instruction-tuned Q&A pairs
- ✓ Considering that perfection is a myth in research, I consider publishing my work.
- ✓ Government: Chatbot integration to support planners & surveyors.
- ✓ RegTech: SaaS product for regulation compliance (subscription model)
- ✓ Innovation: Can be positioned for UKRI Smart Grant as an AI-for-Governance tool

Summary

- ✓ I achieved all objectives: multi-model RAG, model comparison, and domain-specific deployment.
- ✓ LLaMA 3.1-8B delivered best overall performance (80% pipeline correctness)
- ✓ Hybrid retrieval maintained high context recall (0.85) across models.
- ✓ StructureGPT offers a practical, scalable approach to make complex UK Building Regulations more accessible, transparent, and usable by professionals.


AI