Using GenAI to Accelerate Decarbonizing and Electrifying New York City's Real Estate

Final Report | Report Number 26-16 | April 2026

NEW YORK STATE

NYSERDA
New York State Energy Research and Development Authority

NYSERDA's Mission:

NYSERDA catalyzes New York's clean energy transition.

Our Vision:

Clean energy that supports a healthier and thriving future for all New Yorkers.

Our Promise to New Yorkers:

NYSERDA serves New York State as a trusted and credible resource for energy information, policies, and programs, through objective analysis and planning, innovative solutions, and impactful investments that are valued by New York residents and businesses.

Using GenAI to Accelerate Decarbonizing and Electrifying New York City's Real Estate

A Use Case Guide Developed by NYC CRE Owners, Operators, Engineers, and Consultants

Final Report

Prepared for:

New York State Energy Research and Development Authority

Albany, NY

Thomas Yeh

Consultant

Prepared by:

GenAI for NYC Commercial Real Estate Working Groups

New York, NY

NYSERDA Report 26-16 **NYSERDA Contract 258892** **April 2026**

Notice

Preferred Citation

New York State Energy Research and Development Authority (NYSERDA). 2026. "Using GenAI to Accelerate Decarbonizing and Electrifying New York City's Real Estate: A Use Case Guide Developed by NYC CRE Owners, Operators, Engineers, and Consultants, Final Report." NYSERDA Report Number 26-16. GenAI for NYC Commercial Real Estate Working Groups, New York, NY. nyserda.ny.gov/publications.

Abstract

New York State Energy Research and Development Authority (NYSERDA) supported two practitioner working groups, the Owners and Operators Working Group and the Engineers and Consultants Working Group, convened across six structured sessions between October 2025 and February 2026 to identify, evaluate, and prioritize the workflows where generative artificial intelligence (GenAI) can deliver measurable near-term value in NYC commercial real estate. The central finding across both groups is that GenAI's most immediate value lies in shortening the cycles that slow decarbonization decisions.

Keywords

automated compliance checking (ACC), basis of design (BOD), computerized maintenance management system (CMMS), generative artificial intelligence (GenAI), generative pre-trained transformer (GPT), graphics processing unit (GPU), large language model (LLM), large multimodal model (LMM), mechanical schedule generator, multi-agent deep reinforcement learning (MADRL), optical character recognition (OCR), Retrieval-Augmented Generation (RAG) video random access memory (VRAM)

Acknowledgments

This report reflects the collaborative input of two working groups convened to advance the adoption of practical generative artificial intelligence (GenAI) in New York City (NYC) commercial real estate (CRE) decarbonization and electrification workflows. The following participants attended at least two working group meetings and contributed use case submissions, workflow observations, prioritization, and implementation constraints that shaped the report's findings.

Owners and Operators Working Group

Name	Organization
Benjamine Levine	Douglaston Development
Nicholas Salem	The Durst Organization
José Medina-Pardo	The Durst Organization
Joseph House	ESRT
Mary Jiang	Hines
Zach Sussman	Rudin
Eric Chan	Rockrose
Colleen Graham	Tishman Speyer
Colton Curtis	Tishman Speyer
Karen Oh	Vornado

Engineers and Consultants Working Group

Name	Organization
Linda Lam	Amazon
Anthony Annunziata	Aramark Engineering Solutions
Stephane Audrin	Arup
Wilson Patinho	Bloomberg LP
Julian Vaiana	CHA Consulting, Inc.
Vaibhav Sahdev	CodeGreen
Thomas Morrisson	EN-POWER GROUP
Stephen Eber	Energy Worknet LLC
Colin Breslin	Enica Engineering
Jeff Reardon	Enica Engineering
Justin Simko	Goldman Copeland
Curt Rohner	Goldman Copeland
Sarah Welton	GRESB Foundation
Maruti Anumulapalli	ICF

Maxime De Scheemaeker	JUUNOO Inc.
George Martin	Loring Consulting Engineers
Elias Dagher	Principal Engineer
Robert Abdallah	SGH
Brandon Sheiner	Sheiner Energy Engineering and ASHRAE NYC
Gregori Tayco	TBL Technologies and ASHRAE NYC
Cassandra Leyden	WeWork

The New York State Energy Research and Development Authority (NYSERDA) supported the initiative and provided program leadership.

The NYC Chapter of the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) collaborated with NYSERDA to recruit its NYC-based members to serve on the Engineers and Consultants Working Group. Additionally, the Real Estate Board of New York (REBNY) and the NYC Chapter of CoreNet Global helped identify volunteers for both working groups.

Academic advisory teams contributed to prototype development, live experiments, and technical critique, which strengthened the credibility and testability of the working groups' outputs. The Owners and Operators Working Group was supported by Bianca Howard, Ph.D., Assistant Professor, in the Department of Mechanical Engineering at Columbia University, and Ryan Marc Dubois, M.S., Ph.D. candidate, in the Department of Mechanical Engineering at Columbia University. The Engineers and Consultants Working Group was supported by Saptarashmi Bandyopadhyay, Ph.D., Assistant Professor, in the Department of Computer Science at the City College of New York (CCNY) and a team of CCNY students, including Kacper Borowski, Mehedi Hasan, Zhuolin Li, Aarush Kumar, Gilyana Dorzhieva, Patience Divine Aimee Izere, Ahmed Huossein, Axyl Frederick, and Ei Cho.

Thomas Yeh, an independent consultant for NYSERDA, served as moderator across all working group sessions, facilitating discussions, synthesizing work products, and integrating findings into this report.

Table of Contents

Notice ii
Preferred Citation ii
Abstract iii
Keywords iii
Acknowledgments iv
Acronyms and Abbreviations vii
Executive Summary ES-1
1. Overview and Scope 1
2. Data Collection and Interview Approach 2
2.1 Participant Recruitment 2
2.2 Session Structure and Arc 2
2.3 Why New York City? 3
2.4 Methodology and Evidence Standards 4
2.5 Academic Advisory Role 4
3. Use Cases Developed 5
3.1 Owners and Operators Working Group 5
3.1.1 Documentation and Compliance 5
3.1.2 Portfolio Intelligence and Capital Planning 6
3.1.3 Operations and Fault Detection 7
3.1.4 Executive Visibility and Risk 7
3.2 Engineers and Consultants Working Group 8
3.2.1 Compliance and Code Research 8
3.2.2 Documentation and Deliverables 8
3.2.3 Asset Intelligence and Design 9
4. Cross-Cutting Issues 11
5. Prototypes and Academic Perspectives 13
5.1 Columbia University: Owner Workflows and NYC Public Data 13
5.2 City College of New York: Engineering Reliability and Validation Architecture 15
6. Recommendations for Professionals Getting Started 18
6.1 Start with a Workflow Outcome 18
6.2 Establish Data Readiness before Model Tuning 19

6.3 Design Governance as a Workflow 19
6.4 Measure Iteration Reduction 20
6.5 Embed Outputs into Existing Workflows 20
7. Conclusions 21
8. References 23
Appendix A. Use Case Detail A-1
Appendix B. Prototype and Try-It Asset Detail B-1

Acronyms and Abbreviations

ACC	automated compliance checking
AI	artificial intelligence
APIs	application programming interfaces
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
BBL	Borough-Block-Lot
BIM	building information modeling
BIN	Building Identification Number
BMS	building management system
BOD	basis of design
CAD	Computer-Aided Design
CAPEX	capital expenditures
CCNY	City College of New York
CMMS	computerized maintenance management system
CO	certificate of occupancy
CRE	commercial real estate
DEP	New York City Department of Environmental Protection
DOB	New York City Department of Buildings
ECB	Environmental Control Board
ECM	energy conservation measure
FDNY	Fire Department of New York
FEMA	Federal Emergency Management Agency
GenAI	generative artificial intelligence
GFA	gross floor area
GPT	Generative Pre-trained Transformer
GPU	graphics processing unit
HVAC	heating, ventilation, and air conditioning
IDFs	input data files
IDs	identifiers
Image RAG	Image-based Retrieval-Augmented Generation
JSON Schema	JavaScript Object Notation Schena

KPIs	key performance indicators
LL84	Local Law 84
LL87	Local Law 87
LL97	Local Law 97
LLM	large language model
LMM	large multimodal model
MADRL	Multi-Agent Deep Reinforcement Learning
MEP	mechanical, electrical, and plumbing
NOAA	National Oceanic and Atmospheric Administration
NYC	New York City
NYSERDA	New York State Energy Research and Development Authority
O&M	operations and maintenance
OATH	Office of Administrative Trials and Hearings
OCR	optical character recognition
PDF	Portable Document Format
NYC PLUTO	New York City Primary Land Use Tax Lot Output
QC	quality control
QA	quality assurance
Q&A	question-and-answer
RAG	Retrieval-Augmented Generation
REBNY	Real Estate Board of New York
REITs	real estate investment trusts
RFIs	requests for information
ROI	return on investment
ROM	rough order of magnitude
SQL	structured query language
VRAM	video random access memory
VRF	Variable Refrigerant Flow

Executive Summary

New York City's (NYC) commercial real estate (CRE) sector faces an accelerating decarbonization mandate. Local Laws 84, 87, and 97 (LL84, LL87, LL97) have imposed compliance obligations that create a documentation burden, require sophisticated technical analysis, and demand faster, more defensible decision-making at every stage of building ownership and engineering practice. At the same time, the workflows that drive decarbonization, including code and local laws compliance documentation, capital planning, energy auditing, drawing review, and retrofit analysis, remain among the most labor-intensive and iteration-heavy in the industry.

The New York State Energy Research and Development Authority (NYSERDA) supported this initiative to find out whether generative artificial intelligence (GenAI) can change that equation in practice and under what conditions. Two practitioner working groups, the Owners and Operators Working Group and the Engineers and Consultants Working Group, convened across six structured sessions between October 2025 and February 2026 to identify, evaluate, and prioritize the workflows where GenAI can deliver measurable near-term value in NYC CRE. Academic advisory teams from Columbia University and the City College of New York (CCNY) translated the working groups' priorities into working prototypes, testing whether identified opportunities are achievable under real NYC data conditions. The result is a field-tested, practitioner-authored guide to responsible GenAI adoption, grounded in the NYC regulatory environment, data context, and operational constraints that define this market.

ES.1 Working Group Findings

The central finding across both groups is that GenAI's most immediate value lies in shortening the cycles that slow decarbonization decisions. These include fewer review iterations per deliverable, faster time from first draft to final approval, earlier detection of compliance gaps, and a better ability to synthesize unstructured data, including drawings, audit narratives, specifications, and operations and maintenance (O&M) manuals, which have historically been too fragmented to use effectively. These gains are measurable, achievable with tools available today, and most credible when pursued in workflows where the bottleneck is documentation and iterations. New York City's regulatory environment shapes both the opportunity and the risk. The compliance artifacts generated by LL84, LL87, and LL97 create a richer set of publicly available building data than exists in almost any other market. Data that, when handled with the right retrieval architecture and governance, gives NYC real estate organizations a practical advantage in portfolio analysis, due diligence, and capital planning.

The same regulatory environment means that outputs must be defensible: tied to sources, traceable to assumptions, and able to withstand scrutiny from internal stakeholders, clients, and regulatory audiences. Generic artificial intelligence (AI) tools deployed without appropriate safeguards introduce more risk than they eliminate.

The working groups identified 17 use cases across two portfolios. For Owners and Operators, the highest-priority opportunities include AI-assisted drawing review, pre–due diligence risk screening, peer energy use evaluation, portfolio performance comparison, building system condition and remaining useful life estimation, and unstructured document intelligence. For Engineers and Consultants, the leading opportunities include automated code and standards compliance research, AI-assisted engineering documentation, drawing and legacy document interpretation, continuous energy auditing, fault detection and explanation, and basis-of-design (BOD) generation. Appendix A provides the full details for the 17 use cases.

ES.2 Responsible Adoption Requirements

Across both working groups, the following four conditions determined whether a pilot gained traction or stalled:

1. **Data must be scoped before deployment,** defined in terms of what is in, what is out, what is authoritative, and how updates are handled.
2. **Verification must be embedded in the workflow architecture through** retrieval grounding, citation requirements, and defined human review points.
3. **Governance must be designed upfront** roles, approvals, access controls, and escalation paths clarified before outputs start moving through review processes.
4. **Metrics must be tied to workflow outcomes, including** iteration reduction, time-to-deliverable, and compliance exceptions identified earlier, so that pilots produce evidence rather than impressions.

The academic prototypes demonstrated that these conditions are achievable. Columbia’s seven tools showed that NYC public datasets can be made accessible to nontechnical users, that due diligence screening workflows can be automated across fragmented public databases, and that drawing interpretation and code compliance verification are technically feasible with the right retrieval and multi-agent architecture. CCNY’s eight experiments showed that multi-agent verification materially reduces the risk of hallucination in compliance-sensitive engineering tasks, that retrieval grounding is essential for defensible outputs, and that embedded verification reduces rework without requiring wholesale changes to existing workflows.

ES.3 Next Steps for Decision-Makers

The path forward is incremental and deliberate. Start with one high-friction documentation or compliance workflow, define the minimum viable data and governance requirements before the pilot begins, and measure the reduction in iteration over two to four cycles before expanding the scope. The use cases most likely to produce early wins are those in which the primary bottleneck is evidence assembly, document extraction, or the drafting of deliverables that currently require multiple review cycles to reach a usable state.

GenAI will not replace the engineering judgment, owner accountability, or regulatory expertise that New York City's compliance environment demands. Deployed carefully, with defined validation pathways and clear human sign-off roles, GenAI can make the professionals who hold that expertise significantly more effective and meaningfully accelerate the decarbonization work that New York City's built environment urgently requires.

1. Overview and Scope

This report is a practitioner-authored guide to responsible adoption of generative artificial intelligence (GenAI) in New York City (NYC) commercial real estate (CRE). It documents where GenAI can measurably accelerate decarbonization and electrification workflows, and what conditions organizations need to implement before they can adopt it responsibly. GenAI's authority comes from the practitioners who built it: owners, operators, engineers, and consultants working inside New York City's regulatory and operational environment, rather than from vendors, model developers, or generic artificial intelligence (AI) guidance written without reference to how this market actually works.

New York City is the right place to do this work, and a difficult place to do it carelessly. Local Laws 84, 87, and 97 (LL84, LL87, LL97) make compliance documentation a core part of every decarbonization and electrification workflow, with data requirements and reporting artifacts embedded in every project. At the same time, the regulatory environment produces an unusually rich set of publicly available building data that most organizations are not yet using effectively. Tenant constraints, multiparty delivery processes, and professional accountability requirements add further complexity that generic AI guidance rarely accounts for. All of these realities shape the findings in this report.

The report covers five areas in sequence:

1. **Data collection and interview approach:** Describes how the findings were generated and the standards applied to determine whether a use case was credible enough to include.
2. **Use cases:** Presents the 17 workflows that both working groups identified, refined, and prioritized as the most promising near-term opportunities.
3. **Cross-cutting issues:** Maps the systemic blockers and enablers that apply across the full portfolio.
4. **Prototypes and academic perspectives:** Presents the working demonstrations developed by Columbia University and the City College of New York (CCNY) to test whether prioritized use cases are achievable under real conditions.
5. **Recommendations:** Translates all of this into practical starting guidance for organizations ready to run their first pilot.

Readers focused on a specific workflow or role can navigate directly to the use cases most relevant to them; readers evaluating whether to invest in a pilot will find the cross-cutting issues and recommendations sections most useful as a starting point.

2. Data Collection and Interview Approach

Understanding how this report's findings were generated is essential to evaluating their credibility. The evidence base was built in real time through a structured series of practitioner working sessions that served simultaneously as the data collection method and the initiative's primary deliverables. This section describes who participated, how sessions were structured, why New York City's specific context shaped the inquiry, and what standards were applied to determine whether a finding was credible enough to include.

2.1 Participant Recruitment

The New York State Energy Research and Development Authority (NYSERDA) partnered with established industry organizations to ensure that both working groups reflected a credible cross-section of the NYC CRE market. For the Engineers and Consultants Working Group, NYSERDA collaborated with the NYC Chapter of the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) to recruit practitioners actively engaged in building systems design, energy engineering, performance analysis, and supplemented with outreach to NYSERDA-qualified FlexTech contractors experienced in energy audits and retro-commissioning. For the Owners and Operators Working Group, NYSERDA worked through CoreNet Global's NYC Chapter and the Real Estate Board of New York (REBNY), and directly engaged selected real estate investment trusts (REITs) and developers from prior NYSERDA initiatives.

A deliberate inclusion criterion shaped both groups: each participant had already begun their own GenAI exploration and could contribute real, NYC-specific workflow observations. The groups also included both organizations that have committed to GenAI adoption at scale and those testing use cases more selectively, ensuring that the findings reflect the full range of current practitioner postures.

2.2 Session Structure and Arc

The initiative ran from the Owners and Operators kick-off on October 9, 2025, through a final combined meeting on February 19, 2026. Each working group followed a six-meeting arc designed to create traceability from open discussion to concrete work products:

1. **Kick-off:** Establish scope, Chatham House Rule norms, and shared vocabulary; begin surfacing candidate use cases and recurring blockers.
2. **Use case discovery:** Expand and refine use cases; group by workflow stage and data dependency; begin distinguishing pilot-ready from prototype-ready.

3. **Cross-cutting issues:** Formalize the systems view (data, trust/governance, return on investment, or ROI) and align on early validation patterns and workflow integration constraints.
4. **Prioritization and prototyping:** Prioritize use cases and surface the minimum viable data and governance requirements; academic advisory teams begin experiments and prototypes.
5. **Outline validation and next steps:** Review the report outlines for readability, NYC specificity, and practical sequencing.
6. **Final combined meeting:** Owners and Operators and Engineers and Consultants report out to one another and converge on final refinements.

Operating under the Chatham House Rule throughout allowed participants to speak candidly about constraints, failures, and friction, producing a richer evidence base than more formal methods would have permitted.

2.3 Why New York City?

The working groups were convened specifically around NYC practice, and that focus shaped every aspect of how data was collected and evaluated. NYC compliance work under LL84, LL87, and LL97 is core project work, generating documentation, evidence, and reporting obligations that are embedded in every decarbonization and electrification workflow. Owners and engineers are managing risk, deadlines, and auditability alongside technical problems, which changes what useful output means in practice. The real measure of value is faster, more defensible decisions: outputs that are tied to sources, traceable to assumptions, and able to withstand scrutiny from internal, client, and regulatory stakeholders.

New York City also produces an unusually rich set of public datasets and compliance artifacts that were incorporated directly into the working group discussions and prototypes. These include LL84 benchmarking disclosures, LL87 audit findings, and ECM recommendations, LL97 compliance context, New York City Primary Land Use Tax Lot Output (NYC PLUTO) building characteristics, New York City Department of Buildings (DOB) permits and filings, New York City Department of Environmental Protection (DEP) boiler and equipment permitting data, and Borough-Block-Lot (BBL) and Building Identification Number (BIN) identifiers that link datasets across agencies. These datasets provided participants with a common, ground-truthed reference point for evaluating use cases at the portfolio scale.

New York City's operating environment also introduced constraints that shaped which use cases were considered viable. Limited access to tenant spaces frequently determines retrofit sequencing. Multiparty delivery, spanning owners, consultants, contractors, and agencies, means that outputs must be clearly accountable. Professional responsibility for engineering and compliance deliverables remains with licensed practitioners regardless of how outputs are generated. These realities kept governance as a recurring theme across sessions: defining who approves what, on what basis, and through what escalation path.

2.4 Methodology and Evidence Standards

Each use case and cross-cutting theme was evaluated against a consistent set of questions that the working groups applied across every session:

- **Workflow fit:** Where does the task sit in real project delivery or operations, and who acts on the output?
- **Data readiness:** What minimum data is required, and what are the common failure modes, such as missing metadata, version ambiguity, and unit inconsistencies?
- **Governance and trust:** Who is permitted to use what data, who approves outputs, and what is the escalation path when outputs conflict with authoritative sources?
- **Verification pathway:** What checks are feasible and appropriate, including citations, assumption logs, multi-check review, and human sign-off, particularly for engineering-grade deliverables?
- **Measurement strategy:** What key performance indicators (KPIs) demonstrate value in NYC CRE practice, including cycle time reduction, fewer review iterations, fewer requests for information (RFIs) and change orders, and reduced engineering hours per deliverable?

This framework ensured that findings remained grounded in real workflows and that every use case carried a defined path from data input to verified, actionable output. Use cases that could not be evaluated against these criteria were set aside for future development.

2.5 Academic Advisory Role

NYSERDA recruited dedicated academic advisory teams, one per working group, to strengthen technical rigor. Columbia University's Building Energy Research Laboratory (BERL) supported the Owners and Operators Working Group; the CCNY's Department of Computer Science supported the Engineers and Consultants Working Group. Both institutions are members of New York State's Empire AI Consortium. Their role was to translate practitioner questions into testable experiments and demonstrate prototype feasibility in real time, ensuring that use cases were empirically grounded.

3. Use Cases Developed

The use cases presented here emerged directly from the working group sessions described in the previous section. They were surfaced by practitioners, tested against the groups' evidence standards, and refined across multiple sessions. A unifying theme across the full portfolio is that NYC CRE has always produced large volumes of unstructured information, including audit narratives, bid specifications, operations and maintenance (O&M) manuals, scope write-ups, maintenance logs, and other document-based records, and electrification and decarbonization projects are no exception. GenAI changes the equation by enabling teams to incorporate that unstructured material into everyday workflows alongside structured data, without requiring every input to be perfectly normalized first.

The use cases are organized by working group and grouped thematically within each, so that relationships between use cases are visible and readers can identify the clusters most relevant to their own workflows. Appendix A has complete details for each use case, including data sources, expected KPIs, and next steps.

3.1 Owners and Operators Working Group

Across its sessions, the Owners and Operators Working Group consistently returned to two practical realities: that NYC compliance work generates an enormous documentation burden, and that portfolio-scale decision-making is constrained by the difficulty of synthesizing information across buildings, systems, and time. The use cases below reflect both pressures and are grouped around the workflows where GenAI showed the clearest path to measurable impact.

3.1.1 Documentation and Compliance

Compliance-related workflows represent the highest concentration of iteration cycles, evidence assembly burden, and rework risk in NYC CRE. GenAI's strongest near-term contribution in this cluster is reducing the number of passes required to get a document, drawing set, or evidence package to a reviewable state.

AI-assisted drawing review accelerates internal compliance and pre-permit quality assurance (QA) by parsing architectural and mechanical, electrical, and plumbing (MEP) drawing packages for conflicts, missing annotations, and code-compliance gaps before submissions reach DOB or internal review. The tool produces structured checklists of required corrections with plain-language explanations and code citations, reducing resubmissions and shortening time-to-permit.

AI-powered drawing review custom Generative Pre-trained Transformer (GPT) extends this capability for owners who need deeper validation, interpreting drawings against the NYC Building Code, Energy Code, and zoning requirements while performing or validating performance-related calculations such as R-values and system sizing assumptions. It serves as an internal reviewer, accelerating early-stage QA before drawings reach consultants or DOB and reducing time and costs from repeated cycles.

An unstructured document database and chatbot make institutional document intelligence accessible by making historically inaccessible artifacts, such as leases, contracts, audit reports, capital planning records, and O&M manuals, queryable through a secure enterprise interface. The AI tool parses and embeds all document types, extracts structured fields, and enables staff to answer questions about policies, contracts, or historical building information without manual searching, reducing duplicated analyses and repeated document requests across departments.

3.1.2 Portfolio Intelligence and Capital Planning

Portfolio-scale decisions in NYC CRE are frequently delayed by the difficulty of synthesizing building performance data across assets, vintages, and systems. This cluster addresses that gap by bringing together public datasets, operational signals, and unstructured records into tools that support faster, better-informed capital allocation.

AI identification of high and low-performing buildings in a portfolio normalizes and compares multibuilding datasets to rank energy performance, identify drivers of underperformance, and highlight targeted operational or capital improvements. It draws on utility bills, building management system (BMS) time-series data, LL87 audit data, and tenant operating requirements to reduce analytical labor burden and accelerate decarbonization prioritization.

Building system condition and remaining useful life estimation synthesizes open data, audit records, utility patterns, and building metadata to produce condition assessments, probabilistic remaining useful life estimates, and timing for predicted retrofit or replacement. The goal is to improve the accuracy of capital planning and reduce reliance on full audits for initial condition screening.

Portfolio performance diagnostic engine normalizes utility, BMS, audit, and tenant datasets across a portfolio to pinpoint the specific drivers of poor performance, whether operational, system-level, or behavioral, and recommend targeted interventions. It is designed to support portfolio-level capital prioritization and avoid unnecessary audits for high-performing systems.

The decarbonization underwriting tool supports rapid screening during acquisitions by generating probabilistic estimates of existing heating, ventilation, and air conditioning (HVAC), envelope, and mechanical systems from minimal deal-level inputs. It produces a decarbonization pathway, rough capital cost estimates, and a list of critical unknowns that require confirmation during diligence, thereby reducing reliance on consultants for early-stage viability assessments.

3.1.3 Operations and Fault Detection

Operational inefficiencies in NYC commercial buildings frequently go undetected because traditional bill-based tools lack the resolution to surface them. This cluster focuses on GenAI’s ability to bring analytical depth to operational data that owners already have.

AI analysis of time-series energy data combines large language model (LLM) capabilities with anomaly detection to structure monthly and interval utility data, identify spikes, baseload drift, and seasonality abnormalities, and correlate trends with tenant setpoints or operating schedules. It produces narrative explanations of anomalies, reducing the cost of diagnosing each issue and supporting earlier intervention.

3.1.4 Executive Visibility and Risk

Construction and climate risk information in NYC portfolios is frequently dispersed, delayed, or inconsistently assembled. These use cases address the coordination and reporting burden that sits between operational data and executive decision-making.

AI summaries of construction project status and communications connect to project management systems such as Procore, Monday.com, and Asana to automatically summarize project status, highlight blockers and overdue items, and generate tailored reports for various audiences. It reduces time spent gathering status information and increases transparency across business units.

AI-generated climate and resilience risk reports ingest spatial, climate, and building data to generate reports covering flood risk, heat exposure, wildfire smoke, air quality, and insurance implications. Delivered as automated written reports or through an interactive question-and-answer (Q&A) interface, it supports faster, more consistent due diligence during acquisitions and asset management reviews.

3.2 Engineers and Consultants Working Group

The Engineers and Consultants Working Group focused on the places where GenAI could most directly reduce the technical labor burden of decarbonization and electrification work, particularly the documentation, compliance research, and data interpretation tasks that consume significant engineering time. Their use cases are grouped around three workflow clusters.

3.2.1 Compliance and Code Research

Regulatory interpretation is one of the most time-intensive and error-prone aspects of engineering practice in New York City. The use cases in this cluster apply GenAI to reduce research cycles, improve consistency of interpretations, and surface conflicts earlier in the design process.

Automated Code and Standards Compliance Copilot uses a Retrieval-Augmented Generation (RAG) architecture to ingest regulatory documents and respond to engineering queries with relevant clauses, compliance assessments, and flagged ambiguities. Multi-agent workflows cross-check interpretations to reduce the risk of hallucinations. The result is fewer late-stage compliance issues, faster design iteration cycles, and more consistent interpretations across a team.

3.2.2 Documentation and Deliverables

Engineers across both working groups identified documentation as a significant and largely nonanalytical time sink. This cluster addresses the drafting, interpretation, and generation tasks that consume hours without requiring the core engineering judgment that defines professional practice.

AI-assisted engineering documentation and report generation convert raw notes, calculations, and data outputs into structured engineering reports, proposals, audit narratives, and executive summaries. Outputs maintain traceability to source data and assumptions, reducing time to first draft and the number of review cycles required before a deliverable is client-ready.

AI interpretation of engineering drawings and legacy documents applies multimodal GenAI to interpret mechanical, HVAC, and control drawings; extract system relationships, layouts, and design intent; compare drawings to as-built conditions or BMS data; and flag inconsistencies and conflicts. It reduces the weeks of manual review currently required when engineers inherit incomplete or outdated documentation.

The basis of design (BOD) and mechanical schedule generator produces draft BOD documents and mechanical schedules from natural-language project descriptions and structured inputs. It retrieves relevant prior BODs from a portfolio data lake, recommends system configurations based on comparable buildings, and generates a structured draft aligned with internal engineering standards, flagging assumptions and areas requiring engineer review throughout.

3.2.3 Asset Intelligence and Design

This cluster addresses the data integration and design-iteration challenges that sit at the intersection of engineering analysis and capital decision-making. These use cases have greater data dependency, but they address some of the highest-value problems the working group identified.

Automated energy assessment and continuous auditing replace episodic energy audits with an AI-driven continuous auditor that establishes baselines using historical utility data, detects anomalies and inefficiencies, integrates fault detection analytics, and updates findings as new data arrives. It reduces audit costs and improves measurement and verification practices over time.

AI-enabled fault detection, diagnosis, and explanation layers LLM capabilities on top of existing fault detection systems to translate anomalies into plain-language explanations, diagnose likely root causes, link faults to cost, comfort, and energy impacts, and track fault cycles from detection through repair and verification. It reduces troubleshooting time and enables faster corrective action.

Asset analysis "Rosetta Stone" addresses the fundamental data fragmentation problem in building engineering by generating a semantic integration layer that maps asset registers, BMS data, computerized maintenance management system (CMMS) records, O&M manuals, and financial records to a standardized asset model. It extracts asset attributes from unstructured documents, links operational behavior to maintenance history and costs, and produces condition assessments and recommendations.

Design scenario comparison and iteration enable engineering teams to evaluate multiple design options by comparing scenarios across cost, performance, and feasibility dimensions, tracking changes across iterations, and summarizing trade-off implications for both engineers and clients. It reduces redesign effort and supports better-informed decision-making.

Taken together, these 17 use cases represent the practical frontier of GenAI adoption in NYC CRE as defined by the practitioners who work within it. Their credibility comes from being grounded in real constraints: data that actually exists, workflows that actually run, and governance requirements that must actually be satisfied. The next section examines the cross-cutting issues that apply across this entire portfolio: the systemic blockers and enablers that determine whether any individual use case can move from concept to deployed pilot.

4. Cross-Cutting Issues

The use cases in the previous section each carry their own data requirements and implementation dependencies. Across both working groups, four issues surfaced repeatedly as practical blockers that determined whether a pilot gained traction or stalled. Understanding them as a system is more useful than addressing them individually.

Data readiness was rarely about having no data. The real challenge was finding the right inputs, linking them across systems, and knowing which version was authoritative. New York City's public datasets, including LL84 benchmarking disclosures, LL87 audit findings, LL97 compliance artifacts, DOB permits, NYC PLUTO, and BBL/BIN crosswalks, served as valuable anchors for portfolio-scale triage and basic fact-checking. Owner and engineering datasets added the operational layer: drawings, specifications, BMS trend data, utility interval records, CMMS work orders, and capital planning records. The failure modes that derailed AI pilots were consistent: unit and naming mismatches, time-alignment problems, stale or missing metadata, version ambiguity in drawing sets, and fragmentation across owners, consultants, and vendors. The practical implication is straightforward: before tuning a model, define the minimum viable dataset, write down what is authoritative, and document how updates are handled.

Governance surfaced not as a compliance requirement but as a workflow design problem. GenAI outputs become difficult to act on when no one can answer who approves them and on what basis. Both working groups found that unclear document authority, undefined validation steps, and ambiguous approval responsibility produced one of two outcomes: overtrust, which introduced risk, or undertrust, which killed adoption. Even lightweight pilots benefit from a simple approval map covering inputs, checks, reviewers, and sign-off, because without it, teams spend more time debating the credibility of outputs rather than saving time on the underlying task.

ROI was most credibly framed as iteration reduction and faster time to decision. The measurable impact areas that practitioners consistently identified were fewer review cycles per deliverable, shorter time from first draft to final QA-approved package, earlier detection of conflicts and compliance gaps, reduced hours assembling evidence, and fewer RFIs and change cycles downstream. The practical guidance was equally consistent: pick one or two metrics that directly reflect the workflow, such as review iterations per deliverable or hours assembling compliance evidence, and track them from the start. Pilots are easier to sustain when they produce reusable artifacts such as templates, checklists, and review protocols, rather than one-off demonstrations.

Validation was identified as a design requirement. The operative question for most GenAI workflows is whether outputs can be verified efficiently enough to save time while reducing risk. Lightweight but defensible validation approaches included source citations, assumption logs, cross-checks against a second source, defined stop rules for when the tool should defer to a human reviewer, and clear human sign-off workflows. For higher-stakes tasks or ambiguous inputs, running the same task through multiple model runs and reconciling discrepancies before human sign-off, an approach the working groups called multi-model validation, proved effective at flagging uncertainty hotspots before outputs entered downstream workflows. The key condition for early wins is declaring the validation approach up front; an undefined validation process leads either to mistrust and no adoption, or to overtrust and unacceptable risk.

These four issues, data, governance, ROI, and validation, form a governing sequence for any pilot. The next section examines the prototypes and academic experiments that tested these conditions in practice.

5. Prototypes and Academic Perspectives

The use cases developed by both working groups represented practitioner judgment about where GenAI could deliver measurable value. The academic advisory teams were brought in to test that judgment, taking selected use cases from concept to implementation and finding out what actually happens when they are built against real NYC data and real engineering workflows. The two teams approached this from deliberately different angles. Columbia University's Building Energy Research Laboratory focused on owner-facing workflows and NYC public datasets, building tools that demonstrate how nontechnical users can access and act on data that has historically been too large or fragmented to use. CCNY's Department of Computer Science focused on the reliability and validation architecture that engineering workflows demand, testing whether multi-agent systems, retrieval grounding, and embedded verification can produce outputs defensible enough for compliance-sensitive practice. Together, their work addresses the two questions that practitioners most consistently raise: can GenAI actually access and synthesize the data we have, and can we trust what it produces? Appendix B provides complete technical details for each prototype, including architecture descriptions, data sources, and access links.

5.1 Columbia University: Owner Workflows and NYC Public Data

Columbia University's seven prototypes were built around a central observation from the Owners and Operators Working Group: New York City's public datasets are extraordinarily rich, but most building owners cannot use them. Datasets like LL84 benchmarking disclosures, LL87 audit findings, DOB permit histories, and violation records are too large for standard tools, too fragmented across agencies, and too technical for nonspecialist staff to navigate without significant effort. The prototypes tested whether GenAI could close that gap by turning datasets that require specialized software and expertise into tools that any owner or asset manager can query in plain language.

The NYC decarbonization data chatbot was the first demonstration of this principle. Using an agentic RAG architecture with hard-coded NYC-specific calculation tools and preverified data retrieval pathways, the tool allows users to ask complex questions about building stock, energy conservation measures, and LL97 compliance penalties without writing code. Critically, the team avoided allowing the model to estimate regulatory thresholds or carbon coefficients directly. Those values were hard-coded to prevent hallucination on the numbers that matter most for compliance decisions.

The pre–due diligence risk assessment tool took this further by automating an entire screening workflow. For any NYC building identified by its BBL, the tool pulls records from seven distinct public datasets, including MapPLUTO, LL84, LL87, DOB Permits, DOB Violations, ECB/OATH Violations, and Tenant Complaints, synthesizes them, and generates a preliminary risk report in minutes. The report categorizes buildings as low, medium, or high risk. It identifies specific red flags, including active hazardous violations, recurring mechanical permits suggesting end-of-life systems, and impending LL97 penalty exposure. A process that previously required hours of manual aggregation across siloed databases now takes a few minutes.

The peer energy use evaluator addressed understanding not just how a building performs, but why it performs differently from comparable buildings. The tool identifies a cohort of the 100 most similar peer buildings from the LL87 dataset based on property type, vintage, and square footage, compares the target building's energy end-use breakdown against the peer average, and cross-references permit, violation, and complaint history to identify operational or mechanical deficiencies that correlate with high energy use. It then recommends ECMs with data-backed savings estimates based on what has worked for similar buildings in the cohort.

The remaining four prototypes addressed the challenges of interpreting documents and drawings that owners consistently flagged as practical barriers. The Certificate of Occupancy (CO) Reader demonstrated that even decades-old scanned documents can be made machine-readable through a multistage vision-language pipeline that combines optical character recognition (OCR) extraction with LLM repair and vision-augmented grounding. The Construction Drawing Chatbot tackled the harder problem of making full construction drawing sets queryable, building a custom multiresolution image database that allows the model to retrieve and zoom into specific regions of high-resolution drawings. The Code Compliance Checker extended this into a practical compliance workflow, coordinating three specialist subagents, a Text Expert, a Code Expert, and a Page Expert, to verify code compliance queries against both the drawing set and DOB code simultaneously. The Drafter-Critic-Judge prototype directly addressed trust concerns by demonstrating a multi-agent architecture in which one model drafts a response, a second fact-checks it using a different model family, and a third rewrites it while incorporating the critique.

Across all seven prototypes, the consistent finding was that GenAI can be deployed responsibly in owner-facing compliance and portfolio workflows, provided that regulatory values are hard-coded rather than inferred, retrieval is grounded in verified data pathways, and validation is built into the architecture rather than added afterward. Appendix B has complete technical details for all seven Columbia prototypes.

5.2 City College of New York: Engineering Reliability and Validation Architecture

Where Columbia's prototypes focused on making data accessible, CCNY's experiments focused on making outputs trustworthy. The Engineers and Consultants Working Group had been consistent on this point: the barrier to adoption in engineering practice is reliability, and reliability in compliance-sensitive contexts means outputs that can be checked, traced, and defended. CCNY's eight experiments tested whether specific architectural choices, including multi-agent deliberation, retrieval grounding, embedded verification, and fine-tuning, to see if they could produce that level of reliability in practice.

The Multi-Agent Review Board tested whether structured deliberation between multiple independent agents improves output reliability compared to a single model. Using DOB NOW permit classifications as the test case, the team found that multi-agent consensus improved classification accuracy, reduced off-schema outputs, and decreased variance in ambiguous permit descriptions, at the cost of meaningfully higher inference time and graphics processing unit (GPU) memory requirements. The practical implication is that reliability gains are achievable through architectural design, but infrastructure costs must be evaluated alongside performance before deploying at scale.

The Multi-Agent LL87 Audit and Retrofit Planning Tool applied this principle to the workflow where hallucination risk carries the highest consequences: energy-savings estimates used to justify capital investment. A dual-agent architecture paired an Engineer Agent responsible for interpreting schedules, identifying ECMs, and calculating projected savings with a Verifier Agent that independently recalculated savings, cross-checked schedule assumptions, validated utility rate inputs, and flagged inconsistencies between narrative and numerical outputs. In controlled tests, the Verifier Agent caught misinterpretation of partial occupancy schedules, incorrect application of utility rate tiers, and overstated savings due to baseline misalignment.

The Compliance Copilot addressed the code research bottleneck that engineers identified as one of their most time-consuming tasks. Built on a RAG architecture with retrieval, structured response generation with citations, and outputs strictly bound to retrieved text segments, the system returned relevant regulatory excerpts, citation metadata, and suggested compliance logic. The result was reduced time locating relevant code sections, improved traceability, and lower ambiguity in interpretation workflows.

The Automated EnergyPlus Model Generator tested whether GenAI could reliably produce valid EnergyPlus models from plain-language building descriptions, converting inputs such as a three-story rectangular office of 20,000 square feet per floor into valid EnergyPlus input data files (IDFs). The system produced valid IDF syntax in the majority of test cases and significantly reduced manual geometry encoding time, with errors primarily occurring in complex nonrectilinear geometries.

The remaining four prototypes addressed operational intelligence, portfolio screening, and the communication gaps that slow capital decision-making. The Portfolio Electrification Risk Agent integrated DOB permit histories and equipment installation records to identify electrification assets and operational risk signals across portfolios. ThermalComfortBot combined BMS sensor logs, room metadata, and tenant complaint records to generate root-cause hypotheses and suggested corrective actions. The Senior Engineer Digital Twin synthesized historical maintenance logs and compliance indicators to provide contextual troubleshooting recommendations, preserving institutional knowledge. The Engineering-to-Finance Translator converted technical retrofit scope descriptions and LL97 exposure data into executive-ready summaries, reducing the need for iteration between engineering and finance teams.

Across all eight experiments, CCNY's findings converged on five conclusions:

1. Structured multi-agent architectures materially reduce hallucination risk.
2. Retrieval grounding imposes compliance reliability.
3. Embedded verification reduces rework and accelerates iterations.
4. Fine-tuning improves specialization but increases infrastructure demand.
5. Explicit governance and validation frameworks improve the safety and reliability of GenAI deployment.

Appendix B has complete technical details for all eight CCNY prototypes.

Together, the Columbia University and CCNY prototypes demonstrate that the use cases identified by practitioners are achievable under real conditions, with real NYC data, real engineering workflows, and real compliance constraints. The consistent thread across both teams' work is that architectural decisions matter as much as the model itself: hard-coded regulatory values, retrieval grounding, embedded verification, and defined human review points are what separate a defensible pilot from a convincing demo. The following section translates these findings into practical guidance for professionals who are ready to get started.

6. Recommendations for Professionals Getting Started

The use cases, cross-cutting issues, and prototypes in the preceding sections converge on a consistent message: the organizations that will get the most from GenAI are those that start with a specific workflow outcome and deliberately build toward it. This section translates that message into practical guidance on how to select a technical approach, what data readiness actually requires, how to design governance without creating bureaucracy, and how to measure whether a pilot is worth keeping.

6.1 Start with a Workflow Outcome

The working groups' most consistent practical advice was to resist the temptation to begin by deploying a general-purpose tool. The more productive starting point is a specific deliverable, decision, or review step that currently consumes disproportionate time or produces inconsistent results. Once that outcome is defined, the lightest technical approach that can produce it with traceability becomes clear. Four reference architectures cover the majority of NYC CRE use cases:

RAG is the right choice when outputs need to be grounded in a defined document corpus: document Q&A, evidence assembly, requirement extraction, and drafting with citations all fit this pattern. NYC CRE examples include drawing and specification Q&A, assembling evidence packages for LL97, LL84, and LL87, and summarizing audit and ECM narratives. Good output looks like citations with excerpts and a clear statement of what corpus was searched.

Data pipelines with templated summaries are the right choice for repeatable reporting where inputs are structured or can be made structured, such as portfolio triage, benchmarking narratives, and continuous audit summaries. Good output is reproducible, with stable join k eys, time-aligned data, and consistent baseline definitions.

Tool-driven multistep workflows are appropriate for bounded procedural tasks that require chaining steps such as retrieve, extract, compute, draft, and quality control (QC), and are action-constrained. These require strict permissions, logging, defined stop rules, and a review gate before outputs leave the system. Avoid this approach when tool permissions are unclear, data is sensitive, or failure modes are difficult to detect.

Human-in-the-Loop Review is essential for engineering-grade outputs. For compliance and engineering artifacts, GenAI should prepare drafts for human review, with a named reviewer role, defined acceptance criteria, and an audit trail.

6.2 Establish Data Readiness before Model Tuning

Data readiness is a set of minimum viable requirements that vary by use case, and defining it upfront consistently saves more time than any amount of model tuning. For document-heavy workflows involving drawings, specifications, O&M manuals, and audit submittals, the minimum viable starting point is a defined document set with version identifiers, stable BBL/BIN and internal identifiers, and basic metadata covering dates, discipline, system, and file type. For time-series and operations workflows, the minimum is utility billing or interval data with clarified units and time zones, a BMS point list with a targeted trend set, and a sample of CMMS work orders or fault logs. For portfolio triage and prioritization, the minimum is a set of portfolio attributes covering size, use, vintage, key systems, benchmarking history, and key compliance markers, along with a small, well-defined set of decision outputs, such as top retrofit candidates with rationale. In all cases, the practical step is to write it down: what is in scope, what is out of scope, what is authoritative, and how updates are handled.

6.3 Design Governance as a Workflow

Governance surfaced repeatedly as a workflow design problem. The four things worth defining before any pilot are roles, approvals, access controls, and auditability. For roles, identify who initiates the task, who controls the source corpus and permissions, who validates outputs against acceptance criteria, and who signs off for operational use. For approvals, apply light review for internal drafting tasks and required review with sign-off for compliance, engineering deliverables, and capital planning inputs. For access controls, classify sources by sensitivity, including public, internal, client-confidential, tenant-sensitive, and vendor-confidential, and define what may be shared with external tools versus retained internally, and apply least-privilege access for both tools and users. For auditability, log prompts, sources retrieved, and outputs for reproducibility; define escalation paths when outputs conflict with drawings or field conditions; and build in stop rules that require human review when inputs conflict or confidence is low.

Tenant-related data deserves specific attention in NYC practice. Lease terms, concessions, occupancy schedules, complaints, security plans, and access logs should be treated as high-sensitivity by default. Use aggregated or de-identified representations rather than raw tenant records where tenant constraints drive retrofit feasibility.

6.4 Measure Iteration Reduction

Pilots are easier to sustain when measurement is built in from the start. The most credible KPIs for NYC CRE workflows are time-to-deliverable from first draft to final QA-reviewed package; review iterations per deliverable, hours spent assembling evidence packages; reduction in RFIs and change cycles where measurable; and earlier detection of compliance exceptions in the process. Pick one or two metrics that directly reflect the workflow being piloted, establish a baseline before the pilot begins, and run a lightweight feedback loop: capture failure modes, map fixes to data quality issues or retrieval gaps, and maintain a small set of reference tasks to detect regressions when data or documents change.

6.5 Embed Outputs into Existing Workflows

GenAI delivers the most value when outputs are integrated directly into the tools where people already work, including document review platforms, spreadsheets like Excel, BIM coordination routines, and CMMS and BMS workflows, with a clear handoff to the next tasks. Owners hold portfolio context, constraints, and decision authority; engineers and consultants hold technical methods and professional accountability. GenAI tools are most useful when they tighten that handoff by producing review-ready briefs with cited sources, explicit assumptions, and a defined sign-off step, so neither party needs to recreate context from scratch.

A practical starting sequence: pick one workflow artifact such as a compliance evidence pack, drawing extraction summary, or audit triage brief; define inputs, validation steps, and approver roles; and measure iteration reduction over two to four cycles before expanding scope.

The findings across both working groups point toward the same conclusion: incremental, well-governed pilots embedded in real workflows are the most credible path to adoption. The following section draws together the initiative’s central conclusions and maps the path forward.

7. Conclusions

This initiative set out to answer a practical question: Where can GenAI deliver measurable value in NYC CRE decarbonization and electrification workflows today, and what does responsible adoption actually require? Across six sessions, two practitioner working groups answered that question from inside the work, through use case development, cross-cutting issue mapping, and live prototype demonstrations grounded in real NYC data and real engineering constraints.

The central finding is consistent across both groups: GenAI's near-term value in NYC CRE lies in reducing the friction that slows decarbonization decisions. Fewer review cycles, faster time-to-deliverable, earlier detection of compliance gaps, and better synthesis of the unstructured data that has historically been too difficult to use at scale. The 17 use cases documented here represent the practical frontier of that opportunity as defined by the practitioners who live inside these workflows. The fifteen prototypes developed by Columbia University and CCNY demonstrate that the opportunity is achievable under actual NYC data conditions, with validation mechanisms that make outputs defensible rather than merely plausible.

Responsible adoption requires four things that the working groups consistently returned to:

6. Data must be defined before deployment: scoped with clarity about what is in, what is out, what is authoritative, and how updates are handled.
7. Verification must be embedded in the architecture through retrieval grounding, dual-review agents, citation requirements, and defined human sign-off points.
8. Governance must be designed as a workflow with roles, approvals, access controls, and escalation paths clarified before a pilot begins.
9. Measurement must be tied to workflow outcomes such as iteration reduction, time-to-deliverable, and compliance exceptions caught earlier in the process.

New York City is unusually well-positioned to lead this work. The same regulatory density that creates compliance burden, LL84, LL87, and LL97 among them, also produces the richest set of publicly available building data of any market in the country. When that data is handled with the right governance and retrieval architecture, it becomes a meaningful advantage for organizations willing to build on it carefully.

The use cases and prototypes documented here will evolve. Model capabilities are improving quickly, and workflows that are difficult today will become more straightforward over time. The underlying logic will remain constant: GenAI delivers the most value when it is embedded in real workflows, grounded in real data, and validated by the professionals who remain accountable for the decisions it informs. That is the path this initiative points toward, and the one most likely to produce adoption that is durable, defensible, and worth the effort.

8. References

Columbia University, Building Energy Research Lab (BERL). 2025. "GenAI for Commercial Real Estate WorkShop (NYSERDA). Columbia University, BERL. https://berl.me.columbia.edu/research-projects/gen-ai-commercial-real-estate-workshop-nyserda.

Commercial Real Estate (CRE) Working Group. 2025a. "Agentic Vision Drawing Chat." Hugging Face Spaces. https://huggingface.co/spaces/NYSERDA-CRE-Working-Group/Agentic_vision_drawing_chat.

Commercial Real Estate (CRE) Working Group. 2025b. "Certificate of Occupancy Reader." Hugging Face Spaces. https://huggingface.co/spaces/NYSERDA-CRE-Working-Group/docker_c_of_o_extraction.

Commercial Real Estate (CRE) Working Group. 2025c. "Drafter-Critic-Judge Implementation." Hugging Face Spaces. https://huggingface.co/spaces/NYSERDA-CRE-Working-Group/Drafter_critic_judge.

Commercial Real Estate (CRE) Working Group. 2025d. "Modern Decarbonization Chatbot Demonstration." Hugging Face Spaces. https://huggingface.co/spaces/NYSERDA-CRE-Working-Group/Modern_Decarb_Chatbot.

Commercial Real Estate (CRE) Working Group. 2025e. "Peer Energy Use Evaluator." Hugging Face Spaces. https://huggingface.co/spaces/NYSERDA-CRE-Working-Group/Peer_Energy_Use_Evaluator.

Commercial Real Estate (CRE) Working Group. 2025f. "Risk Assessment." Hugging Face Spaces. https://huggingface.co/spaces/NYSERDA-CRE-Working-Group/Risk_Assessment.

Commercial Real Estate (CRE) Working Group. 2025g. "Updated Code Compliance." Hugging Face Spaces. https://huggingface.co/spaces/NYSERDA-CRE-Working-Group/Updated_code_complaince.

Saptarashmi Lab. 2025a. "Automated Compliance Checking Co-Pilot." Hugging Face Spaces. https://huggingface.co/spaces/saptarashmi-lab/acc-copilot.

Saptarashmi Lab. 2025b. "Multi-Agent Auditor." Hugging Face Spaces. https://huggingface.co/spaces/saptarashmi-lab/multi-agent-auditor.

Saptarashmi Lab. 2025c. "Multi-Agent Review Board." Hugging Face Spaces. https://huggingface.co/spaces/saptarashmi/multi-agent-review-board-local.

Saptarashmi Lab. 2025d. “NYC Real Estate: Electrification and Risk Agent.” Hugging Face Spaces. https://huggingface.co/spaces/saptarashmi-lab/nyc-electrification-risk.

Saptarashmi Lab. 2025e. “NYC Real Estate: Technical-to-Financial Translator.” Hugging Face Spaces. https://huggingface.co/spaces/saptarashmi-lab/nyc-fin-translator.

Saptarashmi Lab. 2025f. “Prompting for Auto-building Modeling.” Hugging Face Spaces. https://huggingface.co/spaces/saptarashmi-lab/auto-bem.

Saptarashmi Lab. 2025g. “Senior Engineer Digital Twin.” Hugging Face Spaces. https://huggingface.co/spaces/saptarashmi-lab/engineer-digital-twin.

Saptarashmi Lab. 2025h. “ThermalComfortBot.” Hugging Face Spaces. https://huggingface.co/spaces/saptarashmi-lab/thermal-comfort-bot.

Appendix A. Use Case Detail

The following profiles provide complete details for each use case presented in the report, organized by working group. Each entry includes the problem addressed, the generative artificial intelligence (GenAI) approach, key data sources, expected return on investment (ROI), key performance indicators (KPIs), and next steps and dependencies.

A.1 Owners and Operators

A.1.1 AI-Assisted Drawing Review

Problem: Building plan reviews, both internal and Department of Buildings (DOB)-facing, are slow and inconsistently executed. Owners struggle to catch errors early due to the complexity of drawing sets and the volume of code references to cross-check.

GenAI Approach: A GenAI vision model interprets architectural and mechanical, electrical, and plumbing (MEP) drawings to run prechecks against New York City (NYC) Building Code and Energy Code, identify missing or contradictory details, verify compliance before submission, and produce a structured checklist of required corrections.

Key Data Sources: Architectural drawings in Portable Document Format (PDF) and Computer-Aided Design (CAD), MEP drawings, NYC Building Code and Energy Code text, and zoning metadata.

Expected ROI and KPIs: Shortened permitting and approval timelines, reduced resubmissions, fewer human review hours, and higher quality submissions to DOB NOW.

Next Steps and Dependencies: Build NYC-specific compliance logic, establish governance for drawing uploads, and validate the accuracy of artificial intelligence (AI) with internal reviewers.

A.1.2 AI-Powered Drawing Review Custom GPT for Code Compliance and Performance Checks

Problem: Owners and operators must review architectural and engineering drawings for code compliance, energy performance calculations, and design alignment. These reviews are slow, inconsistent across reviewers, and require multiple iterations with architects and engineers.

GenAI Approach: A Custom Generative Pre-trained Transformer (GPT) dedicated to automated drawing review, capable of interpreting architectural, MEP drawings; checking for NYC Building Code, Energy Code, and zoning compliance; performing or validating performance-related calculations; identifying missing or contradictory drawing elements; generating AI-assisted redline comments; and providing natural-language explanations with code citations and recommended corrections.

Key Data Sources: Architectural and MEP drawings (PDF, CAD, image exports), NYC Building Code and Energy Code text, zoning maps and property metadata, manufacturer specifications, prior DOB comments, and plan review history.

Expected ROI and KPIs: Significant reduction in internal review time, fewer errors discovered late in design or permitting, higher quality DOB NOW submissions, reduced consultant cost for repeated cycles, and faster turnaround on drawing revisions.

Next Steps and Dependencies: Train Custom GPT on internal drawing samples and past review comments, define a secure workflow for drawing ingestion and retention, validate AI accuracy with internal engineers and code specialists, integrate with internal design management tools, and map which review steps remain fully human-driven.

A.1.3 Unstructured Document Database and Chatbot

Problem: Owners and operators generate large volumes of unstructured documents across acquisitions, asset management, leasing, operations, and capital planning. Information is scattered, poorly labeled, and hard to access, causing delays and manual searching.

GenAI Approach: An AI-enabled ingestion pipeline and document warehouse that automatically parses commercial real estate (CRE)-relevant unstructured files, extracts structured fields such as dates, obligations, dollar amounts, and terms, embeds all documents for semantic search, and exposes a secure enterprise chatbot capable of answering questions about policies, documents, contracts, or historical building information.

Key Data Sources: Asset management documents, vendor and service contracts, financial attachments, acquisition and due diligence documentation, corporate memos, internal notes, inspection reports, and emails.

Expected ROI and KPIs: Major reduction in time spent searching for documents, faster completion of underwriting and internal reviews, higher consistency in referencing historical information, reduction in duplicated analyses and repeated document requests.

Next Steps and Dependencies: Define data governance and retention rules, build secure ingestion pipelines for all document types, establish role-based access controls, and pilot with one department before enterprise rollout.

A.1.4 Portfolio Performance Comparison

Problem: Owners struggle to compare performance across buildings due to inconsistent data, tenant variation, and operational differences.

GenAI Approach: GenAI normalizes and compares multibuilding datasets to rank energy performance, identify drivers of underperformance, highlight targeted improvements, and explain why one building uses more energy than another.

Key Data Sources: Utility bills; building management system (BMS) time-series data; Local Law 87 (LL87) and retro-commissioning audit data, building age, type, and square footage; and tenant operating requirements.

Expected ROI and KPIs: Faster identification of low-performing buildings, improved decarbonization prioritization, and lower analytical labor burden.

Next Steps and Dependencies: Secure and consistent BMS and utility data feeds, and a data governance model for cross-building comparisons.

A.1.5 Building System Condition and Remaining Useful Life Estimation

Problem: Owners lack visibility into system condition without conducting full audits. Baseline equipment health is needed for capital planning and decarbonization strategy.

GenAI Approach: AI synthesizes open data, audit records, utility patterns, and building metadata to produce condition assessments, probabilistic remaining useful life estimates, predicted timing for retrofit or replacement, and early identification of systems likely to fail.

Key Data Sources: Local Law 84 (LL84) and LL87 audit and benchmarking datasets, DOB violation history, NYC Open Data building characteristics, and utility consumption data.

Expected ROI and KPIs: Reduced the need for initial audits, improved capital planning accuracy, and faster decarbonization modeling.

Next Steps and Dependencies: Increase data completeness across the portfolio, validate predictions against real maintenance histories, and standardize system classification across buildings.

A.1.6 Portfolio Performance Diagnostic Engine

Problem: Owners cannot easily explain why performance varies across buildings. Identifying true root causes requires cross-system expertise that is difficult and time-consuming to assemble.

GenAI Approach: A portfolio-wide performance engine that normalizes utility, BMS, audit, and tenant datasets; identifies cross-building patterns; pinpoints drivers of poor performance; and recommends targeted interventions.

Key Data Sources: Utility bills, BMS logs, audit reports, tenant operating demands, and weather-normalized energy models.

Expected ROI and KPIs: Highest-impact decarbonization recommendations, avoided audits for high-performing systems, and portfolio-level capital prioritization.

Next Steps and Dependencies: Clean and well-permissioned multibuilding datasets, governance decisions for cross-building analysis, and validation through pilot deployments.

A.1.7 Decarbonization Underwriting Tool

Problem: Owners must quickly evaluate the costs and benefits of decarbonization when considering acquisitions, but deal volume and limited early building information make this difficult. Traditional due diligence provides this analysis only after a building is under contract, delaying decisions and raising risk.

GenAI Approach: A rapid-assessment tool that ingests minimal deal-level inputs and produces a probabilistic estimate of existing systems, an approximate energy intensity and emissions baseline, a decarbonization pathway tailored to the asset type, rough capital cost estimates, and a list of critical unknowns requiring confirmation during diligence.

Key Data Sources: Basic building metadata, listing materials and offering memoranda, NYC Open Data building attributes, historical decarbonization project examples, and energy use models from comparable buildings.

Expected ROI and KPIs: Faster acquisition screening, improved pricing of assets with decarbonization potential or liabilities, reduced reliance on consultants for early-stage assessments, and fewer surprises during formal due diligence.

Next Steps and Dependencies: Assemble training datasets of past retrofits and costs, validate the rough order of magnitude (ROM) accuracy against completed projects, define allowable inputs given confidentiality rules, and pilot in the acquisitions team for 90 to 120 days.

A.1.8 Time-Series Energy Analysis

Problem: Traditional bill-based energy tools do not capture daily or hourly patterns. Owners often cannot detect anomalies or operational inefficiencies at the level of resolution needed for meaningful intervention.

GenAI Approach: Large language models (LLMs) combined with anomaly detection to structure monthly and interval utility data, identify spikes, baseload drift, and seasonality abnormalities, correlate trends with tenant setpoints or operating schedules, and produce narrative explanations of anomalies.

Key Data Sources: Utility interval or monthly bills, weather data, tenant setpoint requirements, and operational schedules.

Expected ROI and **KPIs:** Early detection of inefficiencies, reduced energy waste, and lower cost-to-diagnose per issue.

Next Steps and **Dependencies:** Standard utility data ingestion pipeline, clarification of access rules for tenant-related data.

A.1.9 Construction Status Summarization

Problem: Construction information is dispersed across multiple business units and project management platforms. Leadership receives delayed or inconsistent updates, slowing decision-making.

GenAI Approach: GenAI connects to project management systems to automatically summarize project status; highlight blockers, dependencies, and overdue items; generate tailored reports for different audiences; and provide a question-and-answer (Q&A) interface for project details.

Key Data Sources: Project schedules, budget updates, internal communications, and contractor updates.

Expected ROI and KPIs: Reduced time gathering project status, faster and more accurate decision-making, and increased transparency across business units.

Next Steps and Dependencies: Define access control rules across project teams and establish authoritative project metadata.

A.1.10 Climate and Resilience Risk Reporting

Problem: Acquisitions and asset management teams must synthesize many datasets when evaluating climate and resilience risks. Manual processes are slow and produce inconsistent outputs.

GenAI Approach: AI ingests spatial, climate, and building data to generate reports covering flood risk, heat and urban heat island exposure, wildfire smoke exposure, air-quality risks, and insurance implications.

Key Data Sources: Federal Emergency Management Agency (FEMA) flood maps; National Oceanic and Atmospheric Administration (NOAA) climate projections, hazards, and resilience maps; and building elevation and envelope data.

Expected ROI and KPIs: Faster due diligence, more consistent risk assessments, and better-informed acquisition decisions.

Next Steps and Dependencies: Consultant validation of model output, climate dataset licensing, and access approvals.

A.2 Engineers and Consultants

A.2.1 Automated Code and Standards Compliance Copilot

Problem: Engineering teams spend significant time manually searching, interpreting, and cross-referencing regulatory documents. The process is slow, inconsistent, and prone to interpretation errors when requirements overlap or conflict.

GenAI Approach: A Retrieval-Augmented Generation (RAG)-based compliance copilot that ingests regulatory documents; retrieves relevant clauses in response to engineering queries; assesses compliance of proposed designs, flags ambiguities, and alternative compliance pathways; and provides citations, assumptions, and reasoning steps. Multi-agent workflows cross-check interpretations to mitigate hallucinations.

Key Data Sources: NYC Energy Code and Building Code, ASHRAE standards, Fire Department of New York (FDNY) regulations, LL87 and Local Law 97 (LL97) documentation, project-specific design assumptions.

Expected ROI and KPIs: Reduced time spent on code research, fewer late-stage compliance issues, improved consistency of interpretations, and faster design iteration cycles.

Next Steps and Dependencies: Curated regulatory document ingestion, validation against real plan-review outcomes, and governance for code versioning and updates.

A.2.2 AI-Assisted Engineering Documentation and Report Generation

Problem: Engineers spend a large share of time producing reports, narratives, proposals, and summaries rather than performing analysis. Documentation often requires multiple revisions due to missing context or inconsistent structure.

GenAI Approach: LLM-based assistants that convert raw notes, calculations, and data outputs into structured engineering reports; draft proposals, audit narratives, and executive summaries; adapt outputs to various audiences; and maintain traceability to source data and assumptions.

Key Data Sources: Engineering calculations, audit notes, energy models, field observations, and prior report templates.

Expected ROI and KPIs: Reduced time to first draft, fewer review cycles, increased consistency across deliverables, and lower administrative overhead.

Next Steps and Dependencies: Standardized report templates, clear rules for human review and sign-off, and secure handling of client data.

A.2.3 AI Interpretation of Engineering Drawings and Legacy Documents

Problem: Engineers often inherit incomplete, outdated, or conflicting drawings and documentation. Understanding system intent, equipment relationships, and historical changes requires weeks of manual review.

GenAI Approach: Multimodal GenAI capable of interpreting mechanical, heating, ventilation, and air conditioning (HVAC), and control drawings; extracting system relationships, layouts, and intent; comparing drawings to as-built conditions or BMS data; flagging inconsistencies and conflicts; and generating structured summaries for downstream analysis.

Key Data Sources: Mechanical and HVAC drawings (PDF, scans), legacy design documents, field photos, and BMS point lists.

Expected ROI and KPIs: Faster onboarding to existing buildings, reduced time spent learning the building, and earlier detection of design or documentation gaps.

Next Steps and Dependencies: Improved accuracy of drawing interpretation models, ground-truth validation with engineers, and integration with asset inventories.

A.2.4 Automated Energy Assessment and Continuous Auditing

Problem: Traditional energy audits are episodic, time-consuming, and quickly outdated. Engineers lack tools for continuous, data-driven assessment of building performance between formal audit cycles.

GenAI Approach: An AI-driven automated energy auditor that establishes baselines using historical utility data, detects anomalies and inefficiencies, integrates fault detection analytics, generates iterative energy assessment reports, and updates findings as new data arrives.

Key Data Sources: Utility bills (monthly and interval), weather data, equipment metadata, and fault detection outputs.

Expected ROI and KPIs: Faster assessments, reduced audit costs, earlier detection of inefficiencies, and improved measurement and verification.

Next Steps and Dependencies: Reliable data ingestion pipelines, alignment with accepted audit standards, and validation against traditional audits.

A.2.5 AI-Enabled Fault Detection, Diagnosis, and Explanation

Problem: Existing fault detection systems identify anomalies but do not clearly explain causes, impacts, or corrective actions. Engineers must manually interpret raw outputs to determine next steps.

GenAI Approach: LLMs layered on top of fault detection systems to translate anomalies into plain-language explanations; diagnose likely root causes; link faults to cost, comfort, and energy impacts; and track fault cycles from detection through diagnosis, repair, and verification.

Key Data Sources: BMS time-series data, fault detection outputs, maintenance logs, and equipment specifications.

Expected ROI and KPIs: Reduced troubleshooting time, faster corrective actions, and improved system reliability.

Next Steps and Dependencies: Clean and tagged BMS datasets, integration with CMMS systems, and safeguards to prevent overconfidence in outputs.

A.2.6 Asset Analysis “Rosetta Stone”

Problem: Engineers spend significant time on root cause analysis before recommending solutions. Asset registers, BMS data, computerized maintenance management system (CMMS) records, and financial data are fragmented across systems and formats.

GenAI Approach: A semantic integration layer that maps disparate datasets to a standardized asset model, extracts asset attributes from unstructured documents, links operational behavior to maintenance history and costs, and produces condition assessments and capital recommendations.

Key Data Sources: Asset registers, operations and maintenance (O&M) manuals, drawings and as-builts, BMS data, CMMS data, and utility and financial records.

Expected ROI and KPIs: Faster engineering due diligence, more accurate condition assessments, and improved capital planning support.

Next Steps and Dependencies: Agreement on asset schema (e.g., Haystack or Brick), pilot data access, and relationship mapping across datasets.

A.2.7 Design Scenario Comparison and Iteration

Problem: Evaluating multiple design options is time-consuming, and tracking cost, performance, and risk implications across revisions is difficult to manage consistently.

GenAI Approach: GenAI tools that compare multiple design scenarios, highlight trade-offs in cost, performance, and feasibility, track changes across iterations, and summarize implications for both engineers and clients.

Key Data Sources: Design models, cost estimates, performance simulations, and revision histories.

Expected ROI and KPIs: Faster design iteration, better-informed decision-making, and reduced redesign effort.

Next Steps and Dependencies: Integration with modeling tools, validation of assumptions, and clear boundaries between AI support and engineering judgment.

A.2.8 Basis of Design and Mechanical Schedule Generator

Problem: Engineering teams repeatedly develop basis-of-design (BOD) documents and mechanical schedules across building types. Customizing standardized formats for smaller urban sites, constrained roofs, or tight schedules is labor-intensive and largely manual.

GenAI Approach: A GenAI-driven tool that accepts natural language prompts describing site conditions and constraints, retrieves relevant prior BODs from a portfolio data lake, recommends system configurations based on comparable buildings, generates a structured draft BOD and mechanical schedule, and flags assumptions and areas requiring engineer review throughout.

Key Data Sources: Existing BOD and specification documents, mechanical schedules from completed projects, portfolio building attributes, internal engineering standards, and preferred system templates.

Expected ROI and KPIs: Reduced time to generate initial BOD drafts, faster progression through early design phases, increased consistency across BODs and schedules, and improved utilization of prior engineering work.

Next Steps and Dependencies: Curate and normalize historical BODs and schedules; define standardized input parameters and prompts; establish governance for acceptable assumptions; pilot with a limited set of building typologies before scaling; and validate outputs against engineer-developed BODs.

Appendix B. Prototype and Try-It Asset Detail

The following provides full technical details for each prototype developed by the Columbia University and City University of New York (CCNY) academic advisory teams.

B.1 Columbia University

The Columbia University advisory team focused on translating owner-identified use cases into working demonstrations that integrate generative artificial intelligence (GenAI) with real building data and compliance workflows. The proof-of-concept prototypes and “try-it” assets presented here illustrate how GenAI can be responsibly embedded into decision-making processes within large, portfolio-driven real estate organizations. All demonstrations are available through the Building Energy Research Lab’s project website (Columbia University 2025).

The seven GenAI try-it assets are:

1. The New York City (NYC) Decarbonization Data Chatbot
2. Risk Assessment Tool
3. Peer Energy Use Evaluator
4. Certificate of Occupancy (CO) Reader
5. Drafter-Critic-Judge Implementation
6. Construction Drawing Chatbot
7. Code Compliance Checker

B.1.1 The NYC Decarbonization Data Chatbot

The initial demonstration was inspired by a goal of encouraging interaction with New York City’s rich datasets and is available as an interactive web-based prototype developed for the project (CRE Working Group 2025d).

B.1.1.1 The Challenge: Data Fragmentation and Data Access

New York City maintains some of the nation’s most comprehensive building energy datasets, including Local Law 84 (LL84) benchmarking and Local Law 87 (LL87) audit data. However, these datasets are often too large for standard tools like Excel and require specialized software to interact with them. This creates a barrier for nontechnical building owners who need to cross-reference audit recommendations with Local Law 97 (LL97) compliance limits to make capital planning decisions.

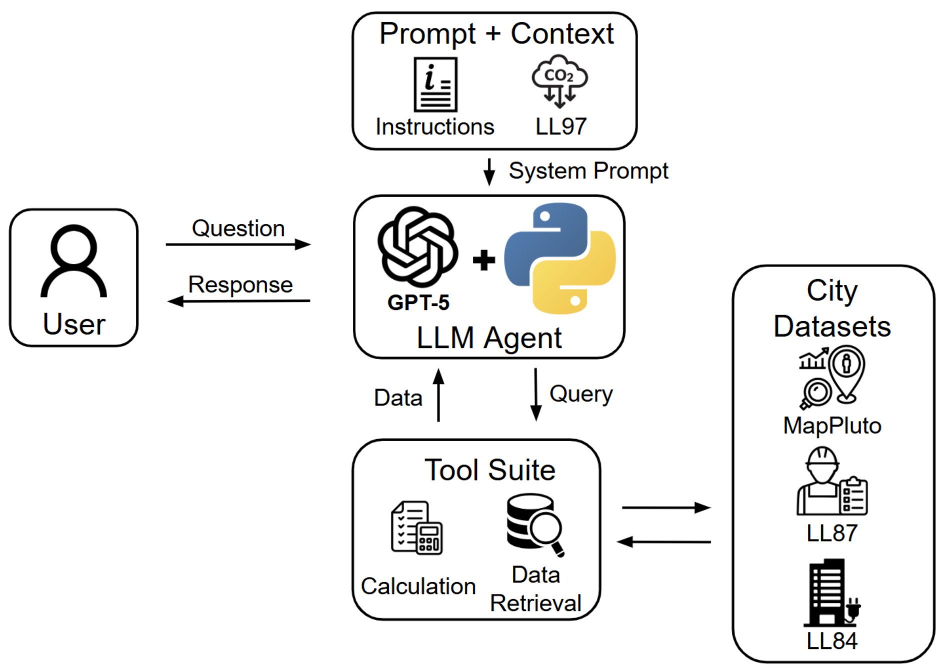
B.1.1.2 This Proof-of-Concept Demonstration: Natural Language as a No-Code Data Interface

This demonstration provides a natural language bridge to New York City's massive building datasets. It allows users to ask complex questions about building stock, energy conservation measures (ECMs), and LL97 penalties without having to write code. This use case is intended to show the value that GenAI workflows can bring to the analysis of large semi structured datasets, highlighting the value of these publicly available datasets.

B.1.1.3 Technical Mechanism: Agentic Retrieval-Augmented Generation with Tool-Calling

Figure B-1 illustrates the overall agentic Retrieval-Augmented Generation (RAG) architecture used in this project. As shown, the large language model (LLM) agent sits between the user interface and a verified tool suite, acting as a reasoning and orchestration layer rather than a direct source of calculations or data.

Figure B-1. Retrieval-Augmented Generation Underlying Architecture



In this framework, user questions are first interpreted by the LLM agent using a system prompt that combines task instructions with NYC-specific contextual guidance (Figure B-1, top). This prompt provides the model with context about LL97 and describes the tool suite available to it, along with general guidelines for answering users' queries.

Once a query is received, the agent enters a multistep tool-calling loop (Figure B-1, center). The agent analyzes the user's intent, selects one or more verified Python tools, executes them, and then synthesizes the returned outputs into a final response.

The tools developed for this project focus on two critical areas to ensure results are grounded in NYC-specific ground truth:

1. **Robust calculation tools:** These functions ensure mathematical consistency by using NYC-defined fuel factors and LL97 emissions limits. By hard-coding these variables, we prevent the model from "estimating" regulatory thresholds or carbon coefficients.
2. **Structured data retrieval:** GenAI frameworks can convert natural language queries directly into structured query language (SQL); however, early implementations had high error rates. To ensure reliability, we implemented a set of predefined data retrieval tasks. These allow the agent to fetch records from LL84, LL87, and MapPluto datasets using robust, preverified pathways identified from common retrieval tasks needed to answer LL97 compliance-based questions. The LlamaIndex natural language query tool is included as a fallback if none of the predefined tools address the model's needs.

B.1.2 Pre–Due Diligence and Risk Assessment Tool

Feedback from the previous demonstration indicated that working group members were, largely, aware of LLMs' ability to serve as chatbots. While they appreciated the ability to interface with these datasets, they were more interested in GenAI to automate workflows. With this in mind, our second demonstration consisted of end-to-end automation of a risk-assessment pipeline, once again utilizing the NYC public databases (CRE Working Group 2025f).

B.1.2.1 The Challenge: Data Fragmentation Barrier to Due Diligence

Although not often used for this purpose, NYC public datasets contain sufficient information for a preliminary due diligence screening of new assets, including complete documentation of the Department of Buildings (DOB) permit, complaint, and violation histories. Using the datasets for this task can require navigating a fragmented landscape of public datasets. Critical "red flags," such as active hazardous violations, history of system failures buried in DOB permits, or large impending LL97 fines, are often stored in separate, siloed databases, such as Permit, Environmental Control Board (ECB), Office of Administrative Trials and Hearings (OATH), LL84, and LL87 datasets. Once again, each of these databases is huge and cannot be opened in traditional tools (such as Excel). Manually aggregating and analyzing this building history for a single Borough-Block-Lot (BBL) is labor-intensive and can take hours to properly investigate.

B.1.2.2 This Proof of Concept: Automated Risk-Screener

This demonstration shows how LLMs can serve as a "Pre–Due Diligence Screener," automatically generating a preliminary Building Risk Report for any specified BBL in a few minutes. The tool acts as a "first pass" for acquisitions or asset management, providing high-level risk scoring by categorizing buildings into low, medium, or high risk, based on available public data for each BBL, and a detailed report covering everything it found about the building that could pose future risks. This assessment includes a summary of the building's permit, compliance, and complaint history, along with a visual of the submitted permits, violations, and complaints.

B.1.2.3 Technical Mechanism: Multi-Dataset Synthesis

Figure B-2 shows the architecture for direct context injection used in this demonstration. Unlike agentic models that "search" for data, this tool performs deterministic aggregation before inference, providing the model with a complete view of the buildings' relevant data. This demonstration highlights the inherent reasoning capability of LLMs rather than relying on complex agentic workflows.

Figure B-2. Multi-Dataset Synthesis Underlying Architecture

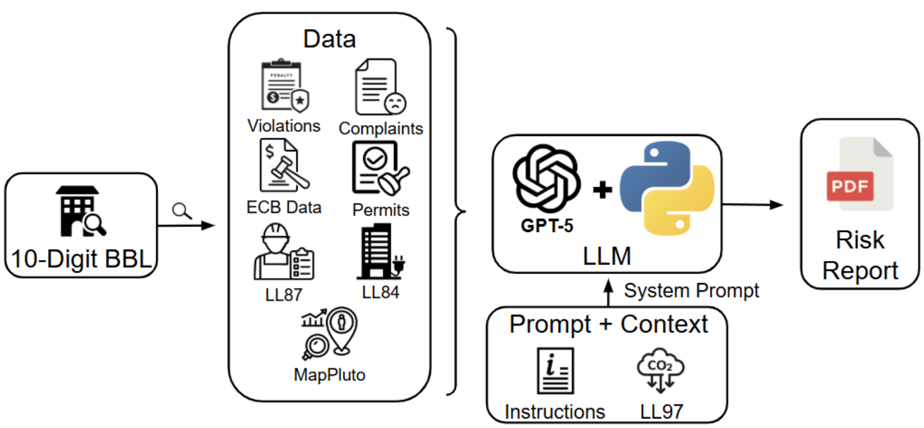


Figure B-2 highlights two key aspects of this workflow:

1. **Data aggregation:** The primary technical mechanism for this project is dataset aggregation. Upon receiving a BBL (Figure B-2, left), the system pulls records from seven distinct ground-truth datasets: MapPluto, LL84, LL87, DOB Permits, DOB Violations, ECB/OATH Violations, and Tenant Complaints. These dataset snippets are passed into the model's context window in full, relying on the model to reason across datasets.

2. **Prompt engineering:** The model is instructed to identify patterns across datasets. These can include checking whether permits were filed to address prior hazardous violations or identifying recurring mechanical permits that suggest a system is nearing its end of life. In addition, the model is provided with explicit, hard-coded regulatory definitions, including the exact LL97 emissions factors and penalty formulas. These minor tweaks to the model serve as safeguards that point it toward a streamlined fact-based analysis.

B.1.3 Peer Energy Use Evaluator

Discussions during the previous working group meeting showed strong interest in leveraging publicly available datasets, particularly permit and violation records, to better understand operational practices within buildings and their relationship to energy use. Building on this idea, we developed a demonstration to compare a building's energy use against that of its peers within the LL87 dataset (CRE Working Group 2025e).

B.1.3.1 The Challenge

Each building in New York City is unique, and its energy use often varies significantly even within the same building typology. Typically, understanding a building's energy use deviations from the norm and giving targeted upgrade recommendations requires a detailed energy audit. Given the vast amount of publicly available information on NYC buildings and the breadth of public audit data, sufficient information may exist to understand a building's operational or mechanical shortcomings by comparing it with its peers.

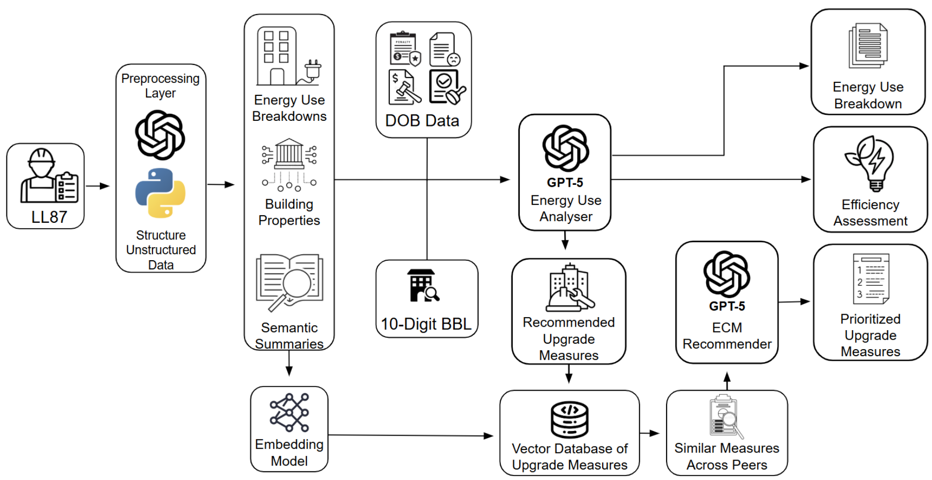
B.1.3.2 This Proof of Concept: Automated Peer Energy Use Comparison

This demonstration uses the LL87 audit dataset, which contains detailed mechanical system data and end-use breakdowns, to contextualize a building's performance. The tool automatically identifies a cohort of the 100 most similar "peer" buildings based on primary property type, vintage (year built), and square footage. It compares the target building's energy end-use breakdown (heating, cooling, and lighting) against the peer average. Then it cross-references the building's permit, violation, and complaint history to identify targeted operational or mechanical deficiencies, such as recurring boiler repairs or elevator complaints, that correlate with high energy use. Using this information along with what worked for similar peers in the cohort, the tool recommends ECMs with data-backed estimations of potential carbon and cost savings.

B.1.3.3 The Technical Mechanism: Vector-Based Peer Matching and Contextual Reasoning

The Peer Energy Use Evaluator uses a more advanced retrieval logic than simple filtering, combining statistical analysis with semantic vector search. Figure B-3 illustrates the full architecture.

Figure B-3: Peer Energy Use Evaluator Underlying Architecture



This architecture can be generally broken down into four components:

1. **Preprocessing:** To transform the semistructured LL87 audit data into a consistent usable format, a preprocessing layer (Figure B-3, left) extracts and parses end-use data by fuel type. All values are normalized into common units to ensure mathematical consistency. Building envelope and mechanical properties are isolated, and ECMs are synthesized into consolidated semantic descriptions. These descriptions are then vectorized using OpenAI's text-embedding-3-small model and stored in a local vector database for high-speed retrieval.
2. **Peer identification:** To ensure a fair comparison between buildings' energy use, the tool calculates the Euclidean distance between the target building and the rest of the NYC building stock across three primary dimensions: Primary Use, Age (Vintage), and Scale, measured as gross floor area (GFA), identifying a relevant cohort of peer buildings.

3. **Curated context injection:** Rather than passing raw data frames to the LLM, the system generates a statistical summary. This compares the target building's fuel usage, end-use profile, and envelope efficiency directly against the peer group's mean and median performance. This structured context is combined with the building's specific permit, complaint, and violation history to give the model the full context about how the target building uses energy differently than its peers, as well as what might be causing those differences (Figure B-3, middle, the Energy Use Analyzer).
4. **Semantic vector database:** A critical technical component of this tool is the vector database of historical audit recommendations (Figure B-3, bottom). Once the Energy Use Analyzer has identified a performance gap and made an ECM recommendation, a semantic search is performed across thousands of peer audit descriptions. This identifies buildings that were recommended similar upgrades and returns their expected normalized energy savings to the ECM Recommender, which summarizes high-confidence recommendations for the user.

B.1.4 Certificate of Occupancy Reader

The CO Reader extracts key data fields from various file formats (CRE Working Group 2025b).

B.1.4.1 The Challenge: Unstructured Data as a Barrier to Entry

Many real estate owners in New York City store their data in various file formats. During previous meetings, specific complaints emerged surrounding decades-old scanned documents that are not machine-readable.

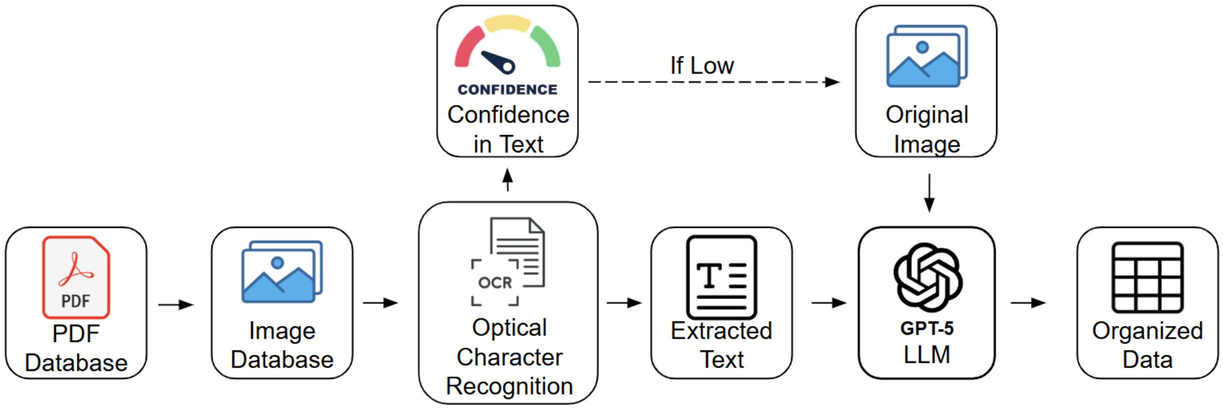
B.1.4.2 This Proof of Concept: Certificate of Occupancy Reader

This simple proof of concept takes any CO Portable Document Format (PDF), in any condition (scanned or digital), and extracts the table from the document, converting it into an organized Excel spreadsheet for the user.

B.1.4.3 Technical Mechanism: Hybrid Multimodal Optical Character Recognition and Semantic Correction

The CO Reader uses a multistage Vision-Language Pipeline, depicted in Figure B-4, that adds a semantic reasoning layer to traditional optical character recognition (OCR).

Figure B-4: Vision-Language Pipeline Underlying Architecture



This high-level workflow can be broken down into four technical components:

1. **Initial extraction:** The system first converts the PDF pages into high-resolution images. It then utilizes Tesseract (OCR) to identify the raw characters and coordinates. This stage captures basic text from newer digital PDFs but often yields messy data with old, scanned files.
2. **LLM repair layer:** To clean up fragmented OCR-extracted text, it is passed to an LLM for correction. The model uses its internal knowledge of NYC building codes and typical CO terminology to repair the extracted text.
3. **Vision-augmented grounding:** In cases where the initial OCR confidence is low (e.g., a smudge on the scan, high noise, or low contrast), the system leverages its multimodal capabilities and requests the original scanned image. The raw image of the page is passed directly to the large multimodal model's (LMM) vision encoder. This allows the model to better reason over low-resolution and noisy scans that are difficult for traditional OCR systems.
4. **Schema enforcement:** To ensure the data is properly structured at output, a final decoding layer extracts the enforced JavaScript Object Notation Schema (JSON Schema) into a data frame for further use by the user.

B.1.5 Drafter-Critic-Judge Implementation

The Drafter-Critic-Judge Implementation approach incorporates various levels of AI in an iterative capacity within the same workflow (CRE Working Group 2025c).

B.1.5.1 The Challenge: Large-Language Model Reliability

Many working group members raised concerns over the trust and reliability of stochastic generative models. One of the demonstrations showcased various techniques to improve the reliability of LLM outputs. Production-ready engineering tools require more reliability than one-shot models can provide.

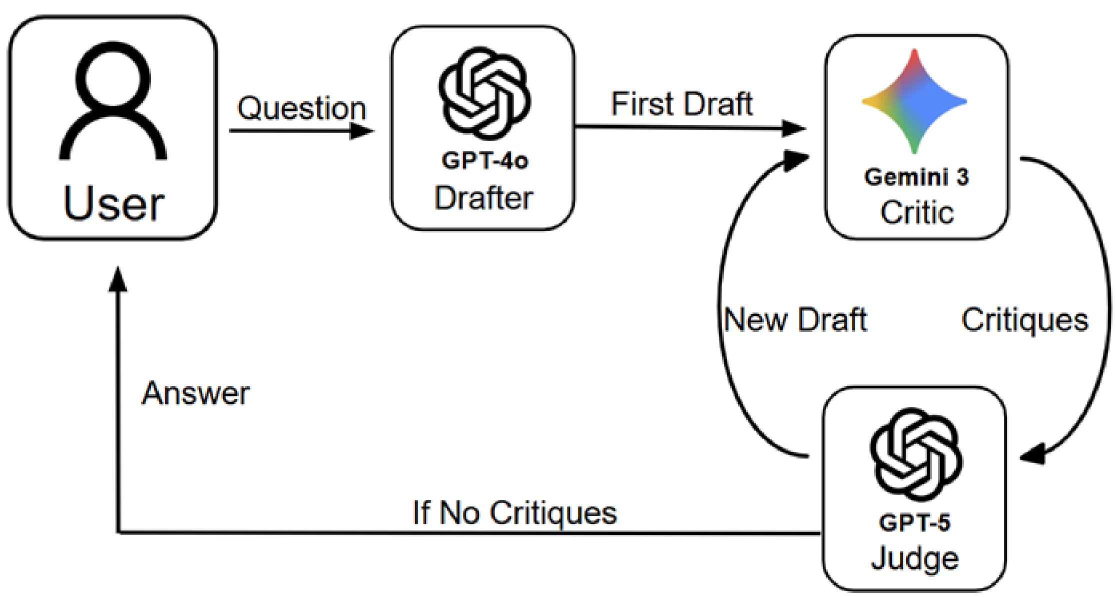
B.1.5.2 This Proof of Concept: Drafter-Critic-Judge

This demonstration showcases a multi-agent architecture designed to improve the reliability of these LLMs, popularized by Andrej Karpathy, a founding member of OpenAI, who argued that agentic self-checking systems must exist as both the creator and the evaluator of their own output. This system is often known as the Drafter-Critic-Judge (or LLM Council) framework, in which one model drafts a response, another model fact-checks it, and a third model rewrites it, incorporating the secondary model's critiques.

B.1.5.3 The Technical Mechanism: Agentic Self-Correction

The power of the Agentic Self-Correction framework, depicted in Figure B-5, lies in model diversity. By using different modes, we increase the probability that others catch a bias in the model.

Figure B-5. Agentic Self-Correction Underlying Architecture



The efficacy of this approach can be broken down into two technical components:

1. **Diverse model selection:** This pipeline uses Generative Pre-trained Transformer (GPT)-4o as the "Drafter" due to its speed, Gemini-3-pro as the "Critic" to provide independent cross-platform verification, and GPT-5 as the "Judge" to perform the final, complex synthesis.
2. **Closed-loop iteration:** In theory, the process is not linear; it is a cycle. If the critic continues to find errors or points for correction in the judge's final response, the loop repeats until both models are satisfied with the result. For this demonstration, the cycle was limited to one forward pass. The judge writes a revised draft and sends it back to the critic for a final critique, but the loop is broken to ensure timely responses for live demonstrations.

Additional quality assurance (QA) methods demonstrated to the working group were Retrieval-Augmented Generation, Chain-of-Thought prompting, and Few-Shot prompting. Only the Drafter-Critic-Judge framework was turned into a functional demonstration for the working group members and is available through the contact information above.

B.1.6 Construction Drawing Chatbot

This demonstration was inspired by a working group member's observation that they were using an enterprise GPT account to accelerate construction drawing reviews. This demonstration is intended to show a more structured workflow that can interact with a construction document for working group members (CRE Working Group 2025a).

B.1.6.1 The Challenge: Construction Drawings as Unstructured Data

Construction drawings are information-dense and critical to many engineering workflows. While modern Computer-Aided Design (CAD) files contain rich metadata, reviewers often work with PDFs or flattened images. Traditional LMMs fail here due to their internal scaling mechanisms. To save compute, these models downsample high-resolution images, reducing the pixel count. In a detailed architectural or structural drawing, this "safe" scaling removes the model's ability to understand fine-line work, callouts, and dimensions. In addition, these documents are often too large to insert directly into the model's context window and must be split into smaller pieces for processing.

B.1.6.2 This Proof of Concept: An Image-based Retrieval-Augmented Generation System for Construction Set Understanding

This demonstration presents an Image-based Retrieval-Augmented Generation (Image RAG) system that enables natural-language interaction with a complete set of construction drawings. The system allows users to ask targeted questions about architectural, structural, and mechanical, electrical, and plumbing (MEP) content and receive answers grounded directly in the visual information contained in the drawings. To better interpret the drawing set, the model has been given the ability to identify relevant sheets and systematically narrow its focus to specific regions of those sheets.

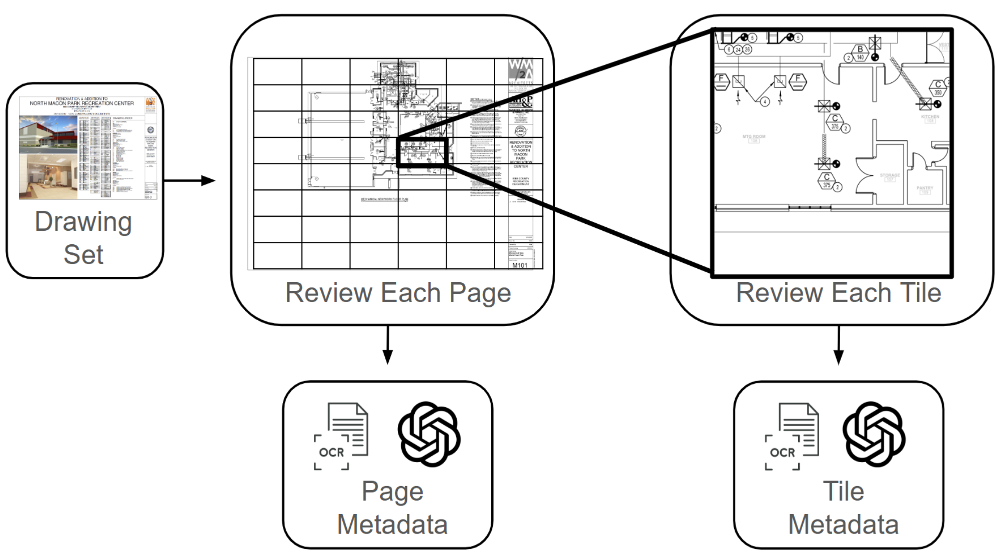
B.1.6.3 The Technical Mechanism: Custom Image Database and Optical Character Recognition

This system employs both a preprocessing pipeline and an inference-time pipeline to overcome the limitations of off-the-shelf LMMs when working with high-resolution, detail-oriented technical drawings. To bypass downscaling limitations, the drawing set is stored as a structured, multiresolution image database, allowing the agent to retrieve and "zoom in" only on the most relevant portions of each page at inference time.

B.1.6.4 Preprocessing Pipeline: Building a Custom Database

Figure B-6 depicts the database construction process.

Figure B-6. Custom Database Underlying Architecture



This process is organized around three primary methodological components:

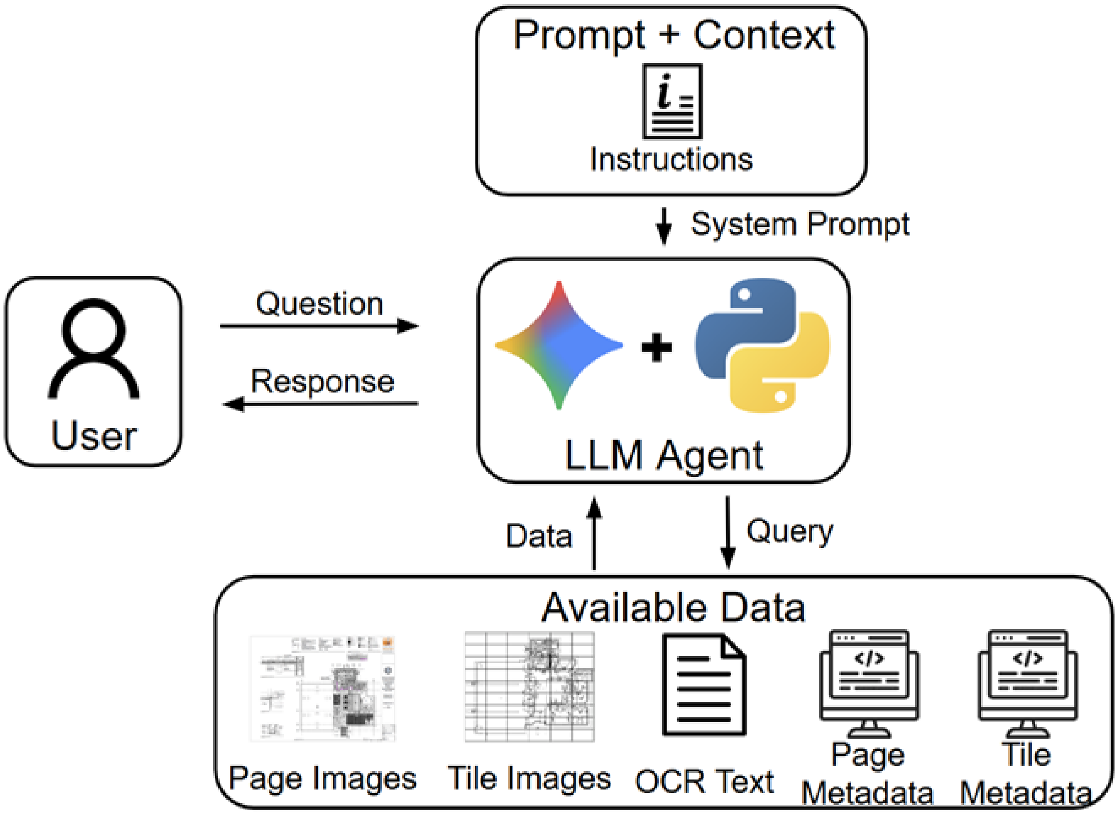
1. **Page-level visual indexing:** Each sheet in the construction set is stored as a full-resolution page image and paired with its corresponding OCR output. An analyst agent processes this combined visual-textual input to generate structured page-level metadata, including sheet name, sheet number, discipline, page description, and the primary views or drawing types present.
2. **Tile-level spatial annotation:** To enable localized visual reasoning, each page is subdivided into 70 spatially indexed tile images. Every tile is tagged with precise coordinate data relative to the parent page and enriched with natural-language metadata describing the visible content within that region: key callouts, symbols, and annotations.
3. **Parallelized metadata generation:** Both page-level and tile-level metadata are generated through parallelized analyst calls to a GPT-based agent. For each page, the full-resolution image and its OCR text are submitted together to produce the page-level metadata. The same process is applied at the tile level, where each tile image is evaluated in the context of its parent page's OCR output to ensure that local annotations and symbols are interpreted within their broader drawing context.

For a 50-page construction set, this preprocessing pipeline completes in approximately 30 minutes.

B.1.6.5 Inference Pipeline: Image Retrieval-Augmented Generation

At inference time, the agent is tasked with the user's natural language query and provided with a system prompt that includes the complete set of page-level metadata, along with access to a suite of custom visual retrieval and image manipulation tools. Figure B-7 shows this Image RAG architecture.

Figure B-7. Image Retrieval-Augmented Generation Underlying Architecture



The inference workflow is organized around two primary methodological components.

1. **Metadata-driven page and tile retrieval:** The agent reasons over the page-level metadata to identify candidate sheets aligned with the user's intent. Once relevant pages are selected, the agent retrieves the full-resolution page image and its associated OCR text, and queries the corresponding tile-level metadata to locate sub-regions of the page most likely to contain the necessary visual information.
2. **Visual context assembly ("zooming"):** After relevant tiles are identified, the agent calls a stitching tool that merges adjacent tiles into a continuous, high-resolution, cropped image. This assembled image allows the model to operate at a scale appropriate for fine-grain interpretation.

Overall, this architecture works similarly to the Custom Image Database and OCR chatbot model. This image database serves as an internal knowledge base that the agent can draw on to construct informed responses.

B.1.7 Code Compliance Checker

Working group members expressed a strong interest in the Chat-with-Your-Drawing workflow and requested an example of this drawing-interpretation methodology applied within a practical workflow. One identified use case was building code compliance verification (CRE Working Group 2025g).

B.1.7.1 The Challenge: Code Compliance as a Time-Intensive Bottleneck

Code compliance verification is a critical preapproval step in the construction review process, requiring practitioners to reason across extensive local code volumes and large construction drawing sets. In practice, this work involves cross-referencing hundreds of pages of regulatory text with dozens of technical sheets to confirm that dimensions, symbols, and spatial relationships satisfy governing requirements and exceptions. This process is highly manual and time-intensive, making compliance review a persistent operational bottleneck for design firms, developers, and reviewing authorities.

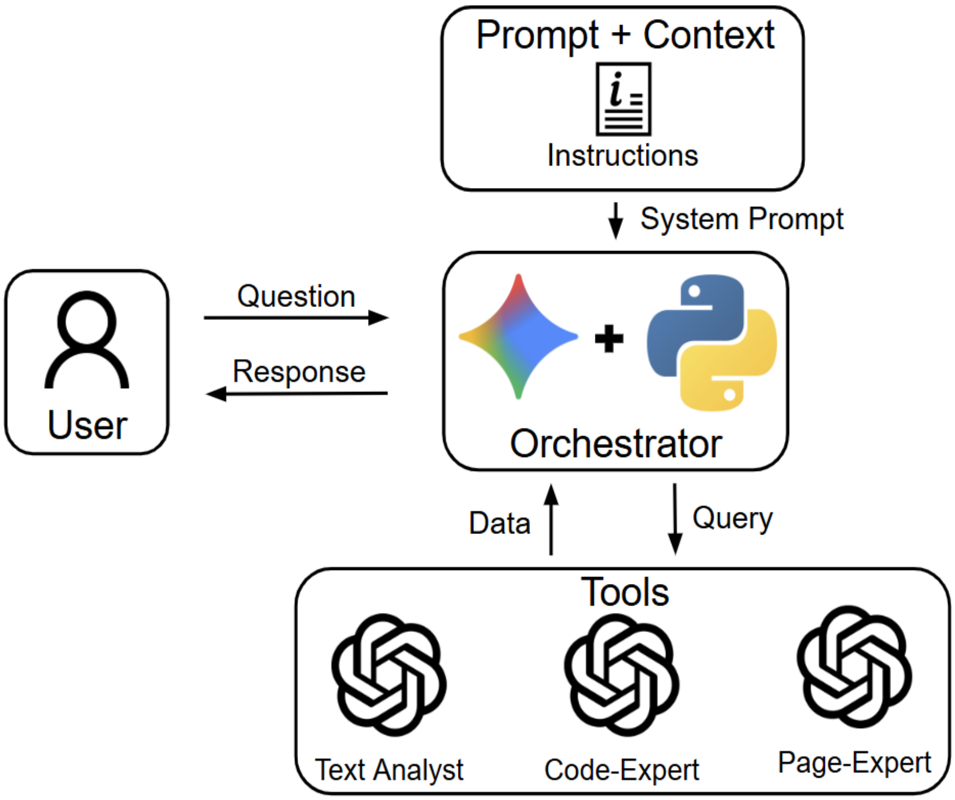
B.1.7.2 This Proof of Concept: Multi–Large Language Model Code Compliance Workflow

This demonstration provides a proof-of-concept workflow that can verify a user's code-related query against legal requirements for a drawing set. It implements a multi-LLM workflow to improve reliability by delegating distinct subtasks to specialist subagents. A central planner agent coordinates three expert agents: a text expert for rapid OCR-based reasoning tasks, a Page Expert for high-resolution visual verification using the Image RAG methodology described in the prior demo, and a code expert for structured retrieval and synthesis of DOB code requirements. This demonstration was designed as a transparent, user-facing workflow. The interface exposes the interaction between agents as it occurs so working group members can observe the workflow in action.

B.1.7.3 The Technical Mechanism: Multi–Large Language Model Orchestration

The technical framework uses a central planning agent (Gemini) that has access to the page-level metadata and is granted access to three tool-based subagents. This is depicted in the same manner as the RAG workflows in Figure B-8.

Figure B-8: Multi–Large Language Model Underlying Architecture



The defining difference between this workflow and previous RAG workflows is that, rather than querying a database, this model queries three subagents as tools that break the problem into bite-sized tasks. The three subagents are described below:

1. **Text expert:** The text expert is designed for fast, low-cost extraction of code-triggering "signals" from page OCR, such as occupancy classification, construction type, and scope of work. Its output is intentionally constrained from interpreting the law or providing compliance conclusions. Instead, it returns short, quote-backed findings that the planner uses to define which code domains and sections are likely applicable before initiating legal research.
2. **Code expert:** The code expert performs legal research over a hierarchical vector database of NYC code (built with ChromaDB). The hierarchy (code volume, chapter, section) is preserved to support structured retrieval. The agent first discovers relevant sections via semantic search, then fetches full chapter context to capture governing rules, exceptions, and cross-references. It returns a structured, citation-backed summary that the planner can directly compare against drawing evidence.

3. **Page expert:** The page expert implements the visual verification methodology from the previous demonstration. It uses page and tile metadata to locate relevant regions, invokes tile stitching to assemble zoomed, continuous image views, and extracts only verifiable claims tied to explicit visual pointers, tile identifiers (IDs), and direct OCR quotes. This agent provides the visual "proof" layer required for compliance verification, especially for questions involving dimensions, symbol interpretation, and spatial relationships.

Interface instrumentation: To support stakeholder trust and improve transparency, the workflow is instrumented so that users can observe agent interaction during execution. Because of this, the agents are called serially, although significant performance gains could be achieved by allowing parallel tool calls. The user interface exposes:

- A real-time tool call log (what agent was invoked and when)
- A subagent analysis panel (intermediate outputs from each expert)
- A code chapter viewer (retrieved legal text used for citations)
- An evidence gallery (full-page context and stitched high-resolution zooms)

This design choice makes the system auditable to working group members and clarifies how GenAI is used in this workflow.

B.1.8 Recommended Future Work

Over the course of this workshop, Columbia University's advisors demonstrated that GenAI can deliver analytical tasks to support building decarbonization and construction management workflows. The tools developed here illustrate an emerging capability to rapidly synthesize heterogeneous data, reason across technical constraints, and accelerate decision-making that has traditionally been labor-intensive. At the same time, these demonstrations reveal substantial opportunities and open research challenges for advancing both the technical foundations and practical deployment of GenAI in the built environment.

- **Building decarbonization analysis with multi-agent workflows:** Through the development of prototypes, we found that a complete building decarbonization analysis requires synthesizing multiple information streams, including historical energy consumption, existing building configurations, available retrofit technologies, and peer comparisons, to assess probable emissions reduction and cost impacts. End-to-end, this process requires Empire-reasoning workflows that integrate data gathering and analytical techniques. Future GenAI-based workflows could incorporate multi-agent systems with access to diverse data collection and analysis tools, enabling comprehensive decarbonization assessments to be completed in minutes. Future work can focus on how best to develop and orchestrate individual agents to align with existing decarbonization and energy efficiency workflows.

- **Spatial reasoning for unstructured building Construction drawings:** Many construction management workflows that GenAI could help streamline require reasoning over building construction drawings that are often in unstructured data formats such as PDFs. Current multimodal models (those with language and vision reasoning) still cannot natively reason over construction drawings, and therefore, enabling reliable spatial reasoning over construction documents remains an open research problem. Potential paths forward include retraining multimodal models for spatial tasks and pairing existing models with geometric tools and domain-specific decoding pipelines.
- **Sustained collaboration between CRE and academia for workflow development:** The workshop has enabled unique direct communication between commercial real estate (CRE) and academia to pair real-world problems with existing GenAI capabilities. Future work should enable these collaborations to continue developing GenAI systems that are not only technically sophisticated but also grounded in real-world decision-making contexts and data infrastructure capabilities, thereby accelerating the translation from experimental prototypes to trusted analytical tools embedded in everyday workflows.

B.2 City College of New York

TheCCNY's Academic Advisory team provided the critical validation architecture required to bridge the gap between theoretical GenAI capabilities and real-world, compliance-sensitive engineering practice. While industry practitioners identified the highest-value priorities, the CCNY team successfully translated these into rigorous, applied experiments and working prototypes. Their vital work focused on proving that generative and agentic artificial intelligence (AI) systems can be highly reliable and technically defensible, ultimately accelerating engineering workflows within the numerically rigorous and highly regulated context of New York City CRE.

Across experiments, the core evaluation criteria were:

- Reduction in hallucination risk
- Improvement in iteration cycles
- Numerical consistency
- Compliance defensibility
- Infrastructure feasibility

The objective was not theoretical performance benchmarking but the practical acceleration of decision-making, iterations, and planning cycles within engineering workflows.

B.2.1 Reliability Architectures for Regulatory and Schema-Sensitive Workflows City College of New York Try-It Assets

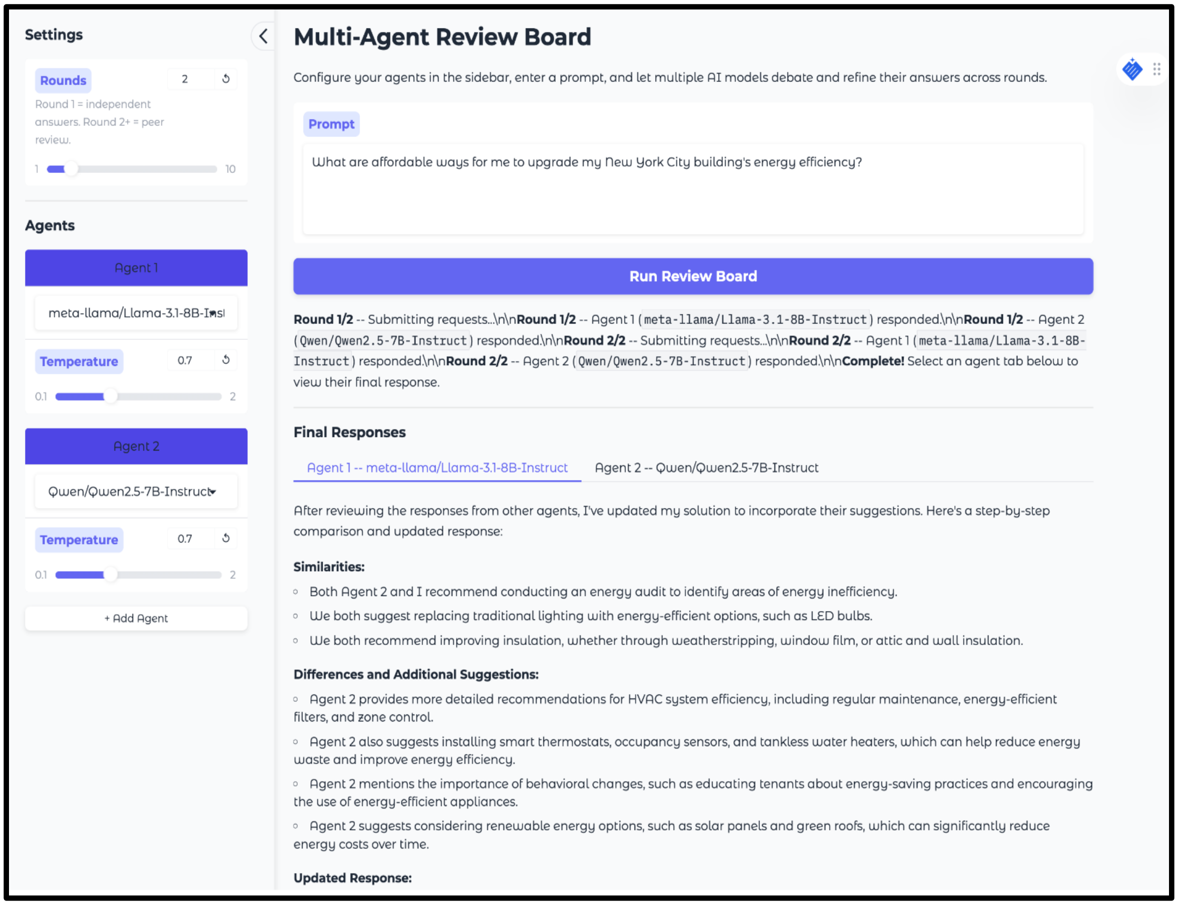
Public-facing prototypes were deployed with Hugging Face Spaces. Access requires a Hugging Face account and an OpenAI Application Programming Interface (API) token, allowing working group members to interact directly with the demonstrations.

1. Multi-agent Review Board
2. Multi-agent Energy Auditor
3. Senior Engineer Digital Twin
4. Automated Compliance Checking (ACC) Co-Pilot
5. NYC Electrification and Risk Management Agent
6. ThermalComfortBot
7. NYC Engineering-to-Finance Translator
8. Prompting for Auto-Building Energy Modeling Tool

B.2.1.1 Multi-Agent Review Board for NYC DOB NOW Permit Classification

To evaluate whether structured deliberation improves output reliability, the CCNY team developed a Multi-Agent Review Board architecture to classify NYC DOB NOW permit job descriptions into official job-type schemas (Saptarashmi Lab 2025c).

Figure B-9. Multi-Agent Review Board



Baseline condition: Single fine-tuned instruction model.

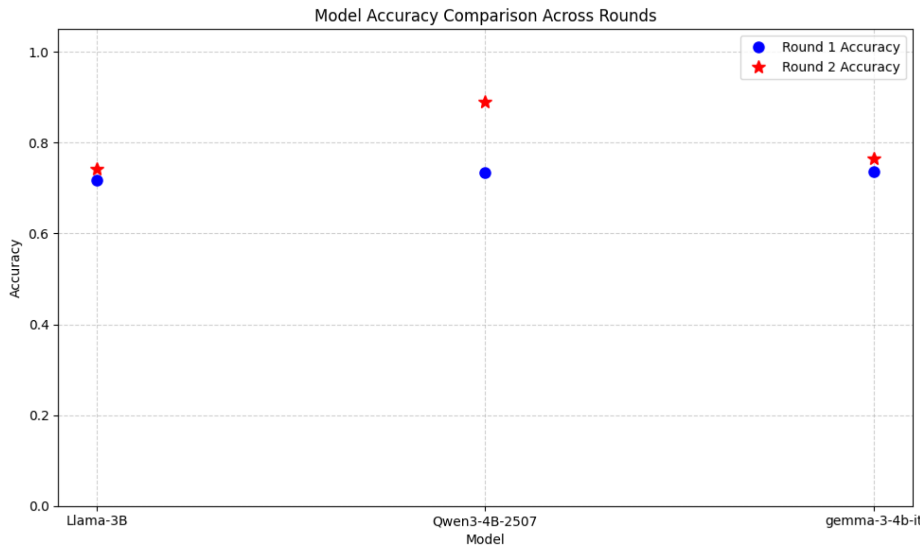
Experimental condition: Multiple independent agents generated intermediate reasoning and proposed classifications. A coordinating arbiter model reconciled outputs through structured consensus logic.

The evaluation used a sampled subset of the NYC DOB Approved Permits dataset. Outputs were normalized to valid job schemas; unmapped outputs were treated as off-schema errors.

Findings

- Multi-agent review improved classification accuracy and reduced off-schema outputs relative to the single-model baseline.
- Structured consensus reduced variance in ambiguous or poorly described permit entries.
- Inference time and graphics processing unit (GPU) memory requirements increased materially under the multi-agent architecture.

Figure B-10. Model Accuracy Comparison across Rounds



Decision impact: Improved schema reliability enables more accurate portfolio-level analysis of electrification permits and capital planning signals, reducing rework and manual correction cycles.

B.2.1.2 Fine-Tuning versus Multi-Agent Deliberation

A controlled comparison examined whether fine-tuning or multi-agent consensus improved task performance more effectively.

Using identical training splits (80/20) and comparable model families (Llama, Qwen, Gemma), the team evaluated:

- Accuracy
- Performance
- Training efficiency
- GPU memory consumption

Observations

- Fine-tuning improved specialization but required significant video random access memory (VRAM) and training time.
- Some multimodal models exhibited unexpectedly high memory demand relative to parameter size.
- Multi-agent consensus improved factual robustness without retraining but increased inference cost.

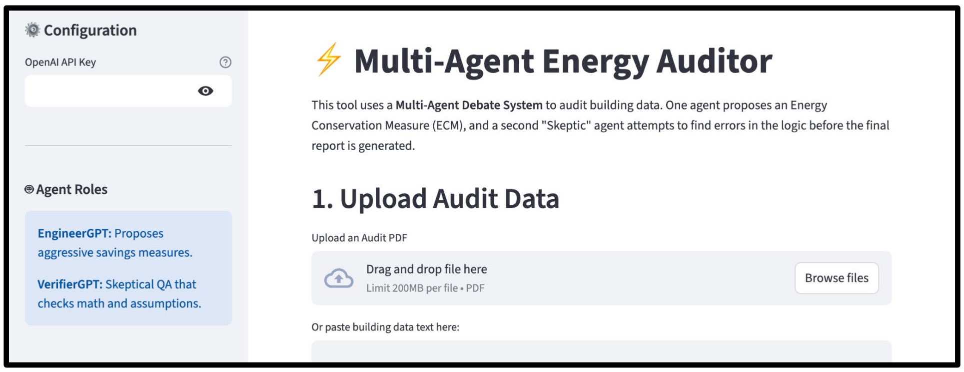
Planning insight: Engineering firms must consider infrastructure constraints when selecting architectures; reliability gains are achievable through design, but scalability must be evaluated alongside performance.

B.2.2 Embedded Verification in Engineering Analysis

B.2.2.1 Multi-Agent Local Law 87 Audit and Retrofit Planning

The Multi-Agency Energy Auditor recommends Energy Conservation Measures (ECM) through an iterative process that incorporates multiple AI levels (Saptarashmi Lab 2025b).

Figure B.11. Multi-Agency Energy Auditor



Try it here: https://huggingface.co/spaces/saptarashmi-lab/multi-agent-auditor

The LL87 experiment focused on reducing logical and mathematical errors in interpreting energy audits and evaluating retrofit measures. Practitioners emphasized that the risk of hallucinations in savings estimates or schedule interpretation can materially distort capital planning decisions. To address this, CCNY implemented a structured dual-agent debate architecture:

Engineer agent responsibilities

- Interpret operating schedules and baseline assumptions
- Identify potential ECMs
- Calculate projected energy savings
- Draft narrative explanations

Verifier agent responsibilities

- Recalculate savings using independent logic
- Cross-check schedule assumptions
- Validate utility rate inputs
- Flag inconsistencies between narrative and numerical outputs

The agents operated in iterative cycles until numerical convergence or logical reconciliation was achieved.

In controlled tests using representative building profiles, the Verifier Agent detected:

- Misinterpretation of partial occupancy schedules
- Incorrect application of utility rate tiers
- Overstated savings due to baseline misalignment

Measured effects

- Reduced numerical discrepancy rates
- Improved logical consistency between narrative and calculations
- Increased review transparency

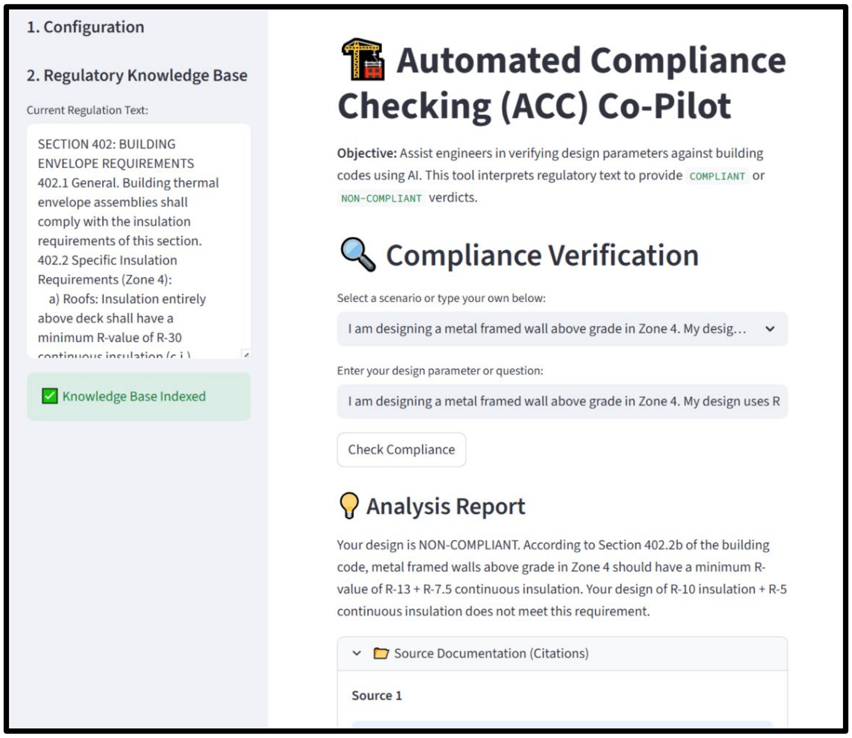
Planning impact: This architecture reduces manual recalculation cycles in audit preparation and strengthens defensibility when presenting retrofit investments for approval.

B.2.3 Compliance Acceleration through Retrieval Grounding

B.2.3.1 Automated Compliance Checking Co-Pilot

The Automated Compliance Checking Co-Pilot assists in interpreting regulatory requirements Saptarashmi Lab 2025a).

Figure B-12. Automated Compliance Checking Co-Pilot



The Compliance Co-Pilot prototype was designed to assist engineers in navigating regulatory frameworks such as American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Standard 90.1 and the International Building Code.

System components

- Document ingestion and segmentation
- Vector embedding generation
- Semantic retrieval via similarity search
- Structured response generation with citations
- Code snippet generation for compliance logic validation

The system returned:

- Relevant regulatory excerpts
- Citation metadata
- Suggested compliance logic blocks
- Confidence scoring

Unlike generic LLM responses, outputs were strictly bound to retrieved text segments to reduce the risk of hallucination.

Measured outcomes

- Reduced time spent locating relevant code sections
- Improved traceability in compliance documentation
- Lowered ambiguity in interpretation workflows

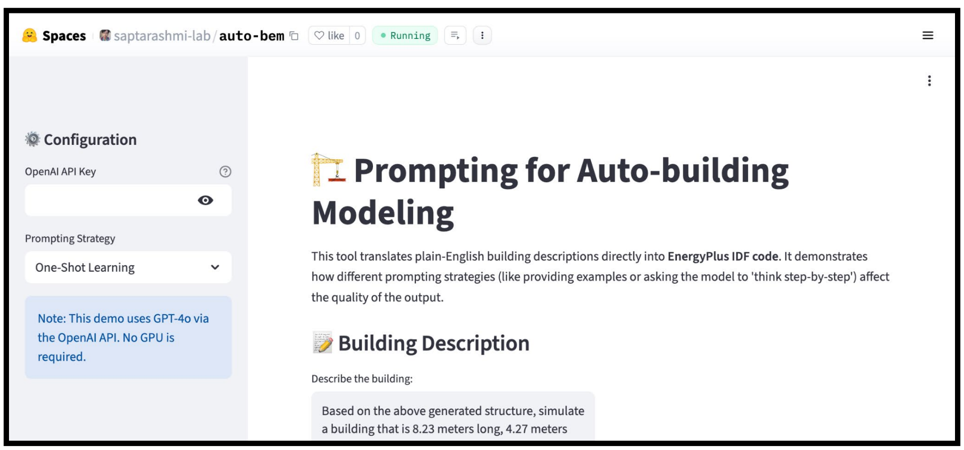
Planning impact: Accelerates compliance review and reduces iteration loops between engineering and regulatory review teams.

B.2.4 Accelerating Early-Stage Modeling and Scenario Planning

B.2.4.1 Automated EnergyPlus Model Generation

The Automated EnergyPlus Model Generation approach translates building descriptions into EnergyPlus code (Saptarashmi Lab 2025f).

Figure B-13. Prompting for Auto-building Modeling



The energy modeling experiment evaluated whether generative AI can reliably produce valid EnergyPlus input data files (IDFs) from natural-language building descriptions.

Key technical features

- Structured prompt templates for geometry parsing
- Stepwise coordinate transformation logic
- Validation checks for surface alignment
- Syntax verification of generated IDF output

  The system converted plain-language descriptions (e.g., “three-story rectangular office, 20,000 square feet per floor”) into:

- ○ Valid geometry coordinates
- ○ Construction assemblies
- ○ Schedule objects
- ○ Heating, ventilation, and air conditioning (HVAC) placeholders

Validation findings

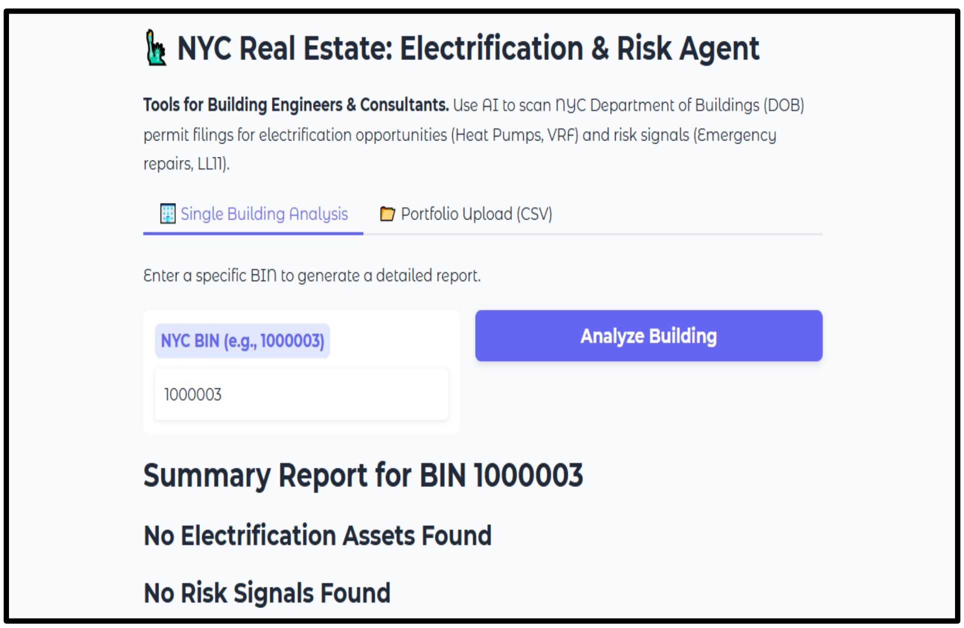
- Valid IDF syntax in the majority of test cases
- Significant reduction in manual geometry encoding time
- Errors primarily occurred in complex non-rectilinear geometries

Planning impact: Enables rapid "what-if" energy modeling during early capital planning discussions.

B.1.4.2 Portfolio-level Electrification and Risk Mapping

The NYC Real Estate: Electrification and Risk Agent analyzes public building info and recommends electrification opportunities (Saptarashmi Lab. 2025d).

Figure B-14. NYC Real Estate: Electrification and Risk Agent



This workflow integrated NYC Open Data application programming interfaces (APIs) with agentic classification models to identify electrification assets and operational risk signals.

Data sources

- NYC DOB permit histories
- Equipment installation descriptions
- Emergency repair filings

Processing steps

1. API retrieval
2. Text embedding and classification
3. Risk tagging
4. Portfolio-level aggregation

The model identified installations of:

- Variable Refrigerant Flow (VRF) systems
- Heat pumps
- Electrified domestic hot water systems

Measured effects

- Reduced manual permit review workload
- Improved visibility into electrification progress
- Enabled early identification of retrofit candidates

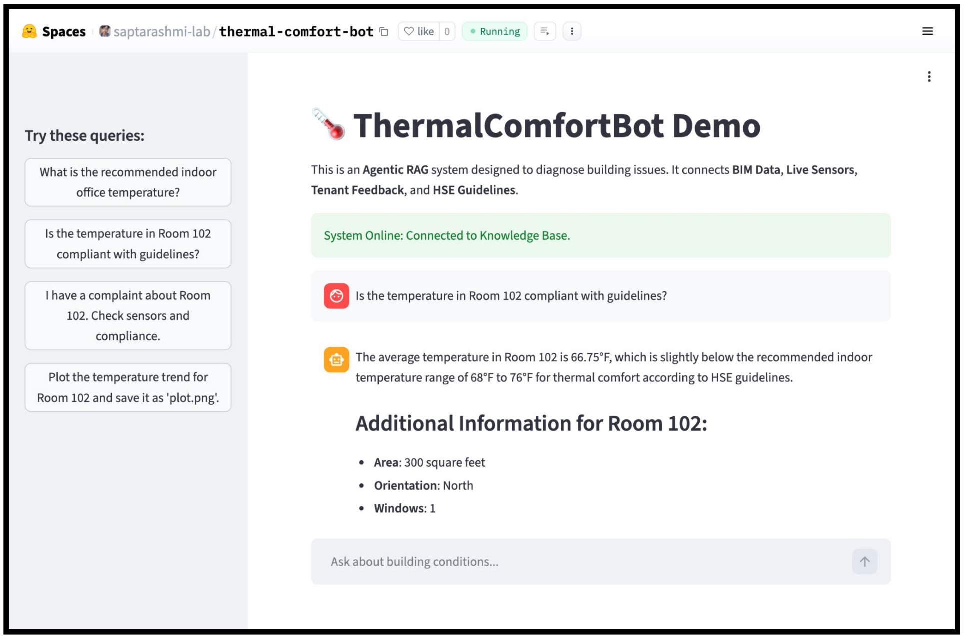
Planning impact: Supports proactive capital allocation decisions across portfolios.

B.2.5 Operational Intelligence and Continuous Commissioning

B.2.5.1 ThermalComfortBot: Real-Time Indoor Environmental Quality Monitoring

ThermalComfortBot is designed to diagnose building issues from existing data (Saptarashmi Lab 2025h).

Figure B-15. ThermalComfortBot Demonstration



This system can connect to data sources such as:

- Building management system (BMS) sensor logs
- Room metadata from building information modeling (BIM)
- Tenant complaint records
- Regulatory comfort thresholds

System architecture

- Statistical anomaly detection
- Retrieval-based regulatory referencing
- Agentic reasoning to correlate multisource inputs

Outputs

- Root-cause hypotheses
- Regulatory context
- Suggested corrective actions

Measured effects

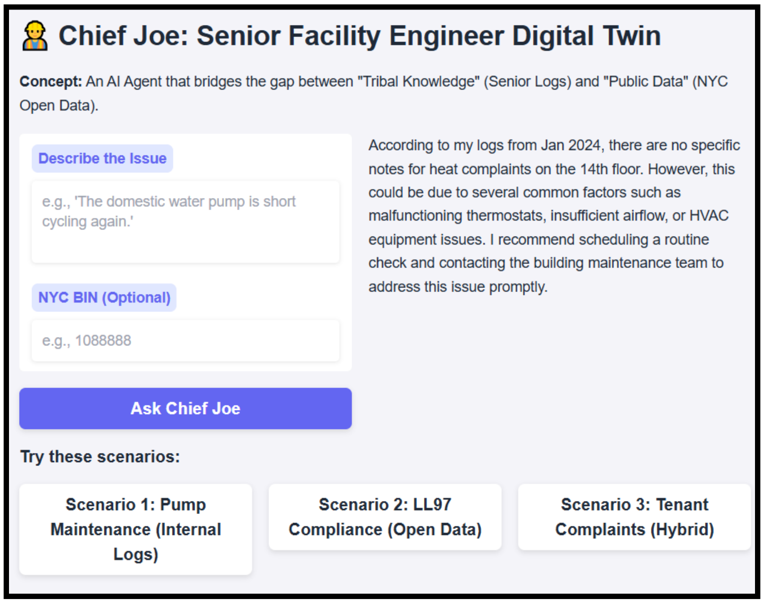
- Reduced time-to-diagnosis
- Improved alignment with ASHRAE comfort standards
- Enhanced transparency in tenant communications

Planning impact: Accelerates operational response cycles and improves compliance alignment.

B.2.5.2 Senior Engineer Digital Twin

The Senior Engineer Digital Twin duplicates a building electronically based on historical data (Saptarashmi Lab 2025g).

Figure B-16. Senior Engineer Digital Twin



The Digital Twin prototype aimed to institutionalize tacit operational knowledge.

Inputs

- Historical maintenance logs
- NYC compliance indicators (LL84 grades, Energy Star metrics)
- Service histories

The system synthesized these inputs to provide contextual troubleshooting recommendations.

Measured effects

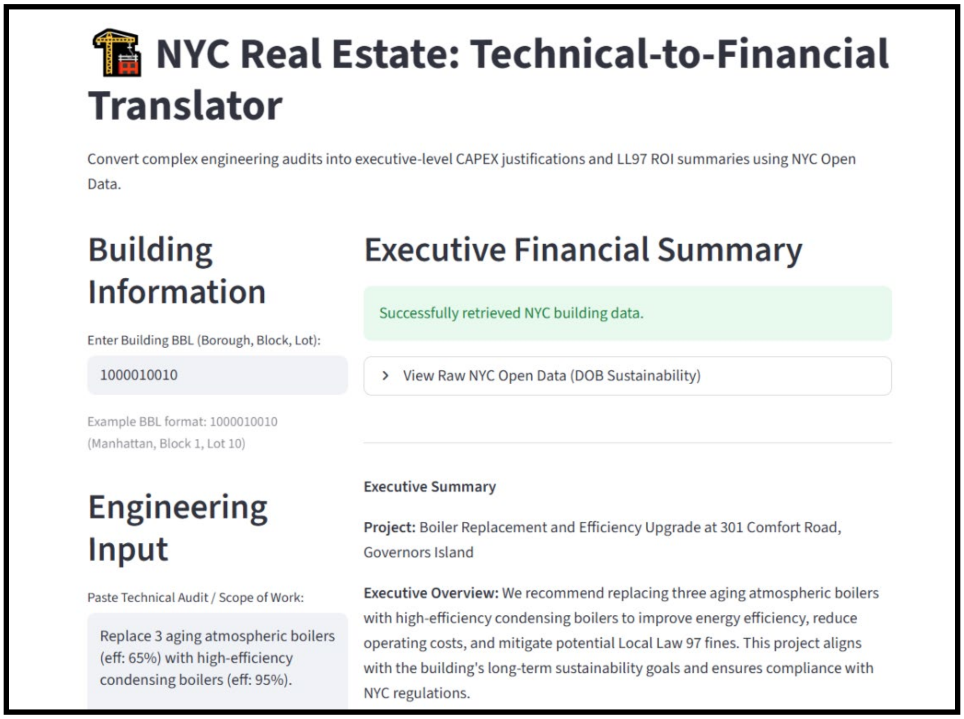
- Reduced reliance on individual engineer memory
- Improved consistency in diagnostic reasoning
- Shortened maintenance decision cycles

Planning impact: Preserves institutional knowledge and supports consistent planning decisions.

B.2.5.3 Engineering-to-Finance Translation

NYC Real Estate: Technical-to-Financial Translator converts engineering audit findings into recommendations based on financial analysis (Saptarashmi Lab 2025e).

Figure B-17. NYC Real Estate: Technical-to-Financial Translator



The Engineering-to-Finance Translator prototype addressed communication gaps between engineering teams and executive stakeholders.

System inputs

- Technical retrofit scope descriptions
- NYC PLUTO building attributes
- Compliance context (e.g., LL97 exposure)

Outputs

- Executive-ready summaries
- Risk framing
- Capital expenditures (CAPEX) contextualization

Measured effects

- Reduced iterations between engineering and finance teams
- Improved clarity in investment discussions
- Accelerated capital planning approvals

B.2.6 Predictive Maintenance and Reinforcement Learning Optimization

To explore continuous optimization potential, CCNY implemented:

- Multi-agent anomaly filtering
- Multi-Agent Deep Reinforcement Learning (MADRL) coordination

MADRL agents optimized HVAC setpoints and equipment coordination under simulated operating conditions.

Findings

- Measurable reductions in simulated energy consumption
- Improved stability in anomaly detection compared to single-model methods
- Increased computational demand

Planning impact: Demonstrates long-term potential for operational efficiency gains while emphasizing infrastructure requirements.

B.2.6.1 Community College of New York's Key Findings on Cross-Cutting Issues

The engineering and consulting working group evaluated the CCNY try-it assets in relation to the identified cross-cutting issues.

1. **Mitigated hallucination risk:** Structured multi-agent architectures materially reduce the risk of hallucination.
2. **Compliance reliability:** Retrieval grounding is essential for compliance reliability.
3. **Accelerated iterations:** Embedded verification reduces rework and accelerates iterations.
4. **Infrastructure trade-offs:** Fine-tuning improves specialization but increases infrastructure demand.
5. **Deployment prerequisites:** Responsible GenAI deployment requires explicit governance and validation design.

Collectively, these experiments demonstrate that GenAI can accelerate engineering analysis, compliance workflows, portfolio screening, and capital planning, provided validation mechanisms and computational constraints are explicitly managed. The findings reinforce the initiative's broader conclusion: effective GenAI adoption in New York City CRE depends as much on architectural rigor, governance clarity, and workflow integration as on model capability.

B.2.6.2 Recommended Next Steps

Advancing Multi-Agent QA/Quality Control (QC) to enhance trust and usability will be critical to drive the adoption of Multi-Agent Review Boards for agentic reasoning in the CRE sector. Future efforts should focus on overcoming core usability and trust barriers to further enhance high-impact use cases. Industry and government professionals require transparent validation protocols to manage professional liability. Future research and development will prioritize producing clear, traceable reasoning for all GenAI agent outputs, enabling engineers to confidently verify the system's logic before implementation.

- **Demonstrating agentic reasoning trustworthiness:** To address stakeholder concerns, going beyond merely showing reduced hallucinations remains a priority. The team will emphasize presenting solid, step-by-step agentic reasoning that strictly complies with the specific, complex NYC regulatory requirements while maintaining fidelity to general enterprise models.
- **Sustainable and affordable engineering within regulations:** Focuses on stakeholder interest in further refining the regulatory framework for real estate and building engineering. GenAI agents facilitate the creation of alternative designs to promote sustainable and cost-effective building practices, while efficiently ensuring regulatory compliance and pinpointing regulations that can be optimized to boost affordability and growth in densely urbanized areas.

In addition to recommending a pilot of the Multi-Agent Review Board in real-world engineering contexts to collect feedback for future deployment, we propose creating a formal framework to evaluate multi-agent reasoning traces for factual accuracy and proper citation of NYC codes. Additionally, we suggest conducting a pilot to explore how a Multi-Agent Review Board might propose potential regulatory reforms.

NYSERDA, a public benefit corporation, offers objective information and analysis, innovative programs, technical expertise, and support to help New Yorkers increase energy efficiency, save money, use renewable energy, and reduce reliance on fossil fuels. NYSERDA professionals work to protect the environment and create clean-energy jobs. NYSERDA has been developing partnerships to advance innovative energy solutions in New York State since 1975.

To learn more about NYSERDA's programs and funding opportunities, visit nyserda.ny.gov or follow us on X, Facebook, YouTube, or Instagram.

New York State Energy Research and Development Authority

17 Columbia Circle
Albany, NY 12203-6399

toll free: 866-NYSERDA
local: 518-862-1090
fax: 518-862-1091

info@nyserda.ny.gov
nyserda.ny.gov

NYSERDA
New York State Energy Research and Development Authority

State of New York
Kathy Hochul, Governor

New York State Energy Research and Development Authority
Charles Bell, Acting Chair | Doreen M. Harris, President and CEO