



Towards Trustworthy Agentic Building Energy Modelling: The Role of Model Context Protocol

Jingfeng Zhou, Meng Wang, Chenxiao Ren, and Rui Tang*

Institute for Environmental Design and Engineering, University College London, London WC1E 6BT, UK

*Correspondence:

Received: July 4, 2026; Accepted: September 16, 2026; Published Online: September 23, 2026; <https://doi.org/10.59717/ipj.energy-use.2026.100064>

© 2026 The Author(s). This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by/4.0/>).

Citation: Zhou J., Wang M., Ren C., et al. (2026). Towards Trustworthy Agentic Building Energy Modelling: The Role of Model Context Protocol. *Energy Use* 2:100064.

Building Energy Modelling (BEM) is beginning to undergo an agentic transition. Recent work has shown that Large Language Models (LLMs) can support model generation, debugging, calibration and retrofit analysis, while the Model Context Protocol (MCP) offers a route for exposing simulation tools as callable and inspectable operations. This Perspective argues that the value of MCP in BEM lies not merely in enabling LLM-based agents to interact with simulation engines, but in allowing agents to call tools, connect fragmented information flows and record the modelling actions through which heterogeneous building information is translated into building performance simulation models. BEM is an information-to-parameter modelling chain, involving geometry, materials, internal loads, systems, controls, weather, operation and calibration. These inputs are rarely available from a single source. They need to be assembled from sources such as Geographic Information Systems (GIS), drawings, Building Information Models (BIM), operational records and semantic metadata. MCP can support this process by providing a protocol layer for coordinating bounded agentic workflows for data extraction and information delivery. It further frames MCP-enabled agentic BEM as a route to traceable information integration, documented model correction and reproducible modelling decisions.

REFRAMING AGENTIC BEM WITH MCP

Building Energy Modelling (BEM) is essential for low-carbon design, retrofit assessment, operational optimisation and energy system integration. Yet its practical challenge does not lie in simulation engines alone. It lies in the conversion of incomplete, scattered and heterogeneous building information into a coherent physical representation. A BEM model is not merely a file that produces energy outputs. It is a structured record of geometry, constructions, internal gains, schedules, heating, ventilation and air conditioning (HVAC) systems, controls, weather, occupancy and associated uncertainties.

Recent studies have begun to apply Large Language Models (LLMs) in building energy applications through natural-language interaction, code assistance, model generation, semantic interpretation, operational analysis and decision support.^{1,2} Their practical effects include lowering the expertise barrier for formulating and inspecting modelling tasks, reducing repetitive work in creating and debugging model objects, and translating unstructured building descriptions into structured modelling actions. EPlus-LLM demonstrated automated EnergyPlus model generation,³ while agentic workflows extended this capability to task decomposition and iterative correction.⁴ In a recent residential testbed, MCP-enabled model inspection, modification and analysis reduced typical task time by 80–90%, from 1–2 h to approximately 9–12 min, at the same time, the average token usage per run is only 47085 with a standard deviation of 7912 tokens.⁵ These results demonstrate benefits for bounded tasks, not fully autonomous information-to-BEM generation.

The Model Context Protocol offers a useful way to address this challenge. MCP is a software protocol layer, not a language model or simulation engine. The LLM interprets intent, plans steps and selects tools, and MCP standardises tool discovery, invocation and contextual exchange. Connected BEM tools execute deterministic operations, and the modeller approves consequential assumptions. In practice, MCP is implemented through a client-server architecture in which an AI application acts as the client and MCP servers expose machine-readable tool schemas. EnergyPlus-MCP is one such open-source server, exposing 35 tools for model management, inspection, editing, simulation and output handling.⁶ Thus, MCP does not itself generate or validate a BEM model; rather, it provides the interface through which bounded, named and logged operations are coordinated. A recent MCP-enabled workflow applied this architecture to conversational and agentic BEM tasks.⁵ Figure 1 summarises this division of responsibilities and the resulting information-to-parameter workflow.

BEM OBJECTS AS AN INFORMATION-TO-PARAMETER CHAIN

A serious discussion of MCP-enabled BEM should begin with the anatomy of the model. Geometry and zoning define building form, floor levels, thermal zones, surfaces, openings, shading objects and adjacency. Envelope inputs define constructions, glazing, thermal properties, thermal mass and infiltration. Internal loads define occupancy, lighting, plug loads, equipment and domestic hot water. Schedules describe when these loads occur. HVAC inputs define plant, air systems, terminal units, ventilation rates, setpoints and control logic. Weather and site context define boundary conditions. Calibration then tests how far the model can explain measured performance.

Each attribute has a different information basis. Drawings may support geometry. BIM models may provide accurate geometric and spatial descriptions of walls, windows, spaces and equipment, yet they often do not contain complete, reliable or simulation-ready thermal properties, operating schedules, control logic or simplified system representations required for energy simulation. Operational data may reveal total energy use but not zone-level behaviour. Specifications may name equipment but not control logic. GIS footprints may support building stock modelling but not a detailed retrofit decision. Therefore, the modeller needs to decide how each type of building information should be converted into BEM inputs, which source should be treated as authoritative, which values are inferred and which assumptions require review. This makes information coordination part of the modelling task itself, rather than a preliminary data-cleaning step.

This is where MCP can support a shift in the field. The aim is not to ask an LLM to fill uncertain or incomplete inputs from its own internal knowledge. The aim is to make the information-to-parameter chain explicit. Each object in the model should carry information about where it came from, how it was transformed, which tool produced it, what uncertainty remains and whether a

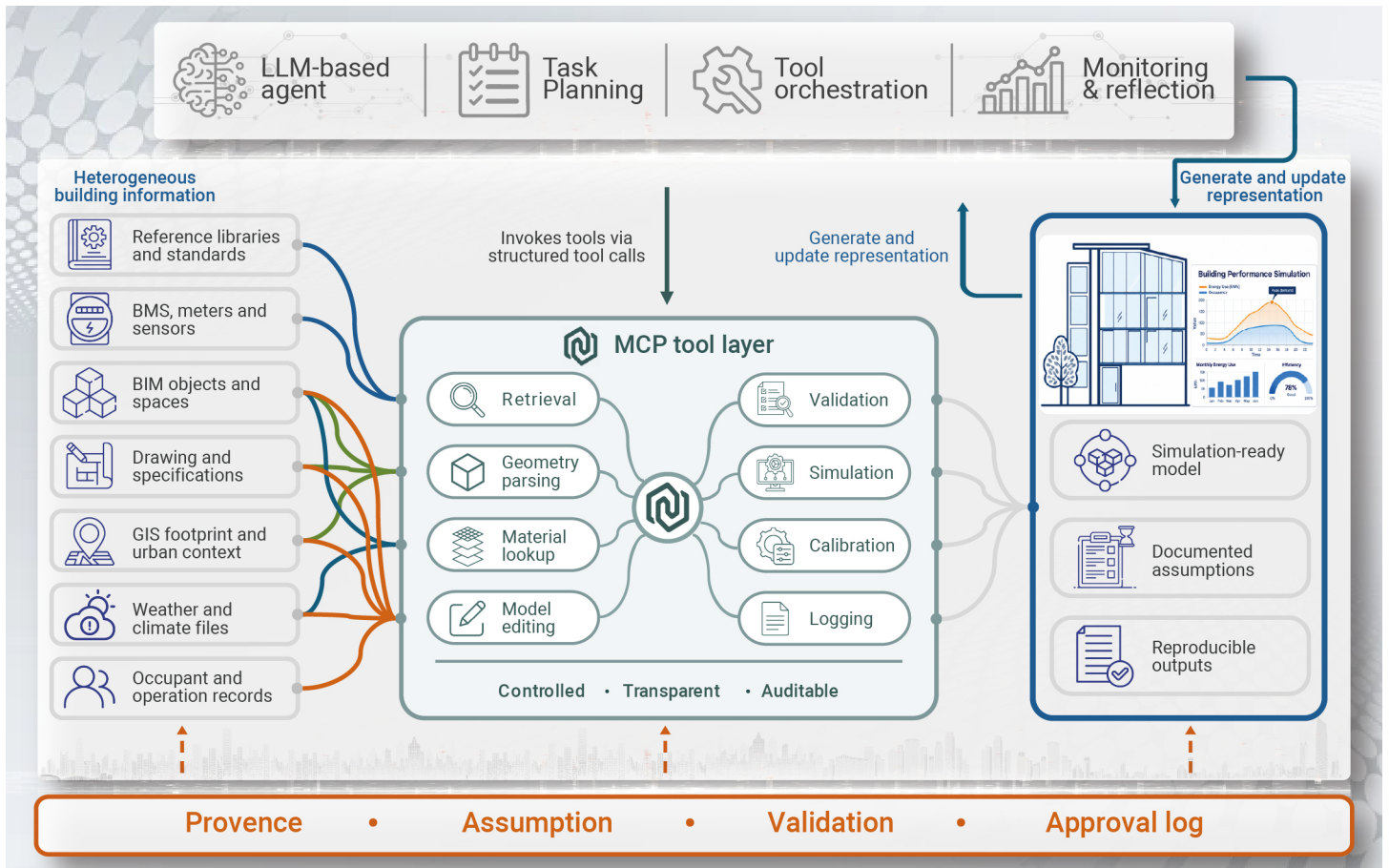


Figure 1. MCP-enabled information-to-parameter modelling for trustworthy agentic BEM.

human approved it. This changes the unit of trust. Trust no longer rests on the fluency of a generated explanation, but on whether the modelling chain can be inspected and tested. In an MCP-enabled workflow, diagnostic tools can review the generated objects, validation routines can check them against schemas and physical constraints, repair tools can propose corrections, and human modellers can approve changes that affect interpretation or decisions.

FROM HETEROGENEOUS INFORMATION TO TRUSTWORTHY BEM GENERATION

A useful long-term framing for MCP-enabled modelling is information-to-BEM generation. Current demonstrations, however, address bounded tasks rather than fully autonomous translation from heterogeneous sources into simulation-ready models.

In GIS-to-BEM generation, footprints, heights, typologies, archetypes, surrounding context and weather data should be converted into urban or portfolio-scale models. This pathway is useful for stock analysis, resilience assessment and district-scale retrofit planning, but it carries archetype uncertainty. In drawing-to-BEM, floor plans, sections and specifications should be translated into thermal zones, openings, constructions and system assumptions. This pathway contains geometric and semantic uncertainty. In BIM-to-BEM, objects created for design coordination need to be simplified, validated and mapped into thermal simulation objects. Recent work on BIM and BEM interoperability shows that data exchange remains difficult because the semantic needs of performance simulation differ from those of design documentation, as well as the data quality of BIM data.⁷ In text-to-BEM, energy audits, design briefs and retrofit reports should be parsed into structured assumptions. In operation-to-BEM, Building Management System (BMS) records, meter data and sensors should be aligned with schedules, setpoints, equipment states and calibration parameters. Recent multi-agent work on automated BEM for energy retrofits has shown how drawings, specifications and sensor data can be combined to support model generation and

calibration, making operation to BEM a key test case for agentic modelling.⁸

MCP allows LLM-based agents to organise these translations as tool-mediated workflows rather than one-step generation. Data connectors exposed as MCP-enabled tools can collect drawings, BIM, GIS, specifications and operational records. Extraction and deterministic preprocessing tools can then standardise formats and units, identify missing or duplicate records, and structure candidate inputs against BEM schemas. An LLM-orchestrated workflow combining retrieval-augmented generation (RAG) with MCP-enabled tools has been applied to urban-scale simulation-based resilience assessment,⁹ which illustrates how MCP can support the coordinated transformation of heterogeneous information into structured simulation inputs. The output should be a modelling information package containing the source map, structured inputs, generated model, object-level provenance, simulation outputs, reproducible visualisations and an action log. Post-processing tools exposed through MCP can generate tables, load profiles, geometry views and comparison plots, while the LLM provides a separately reviewable interpretation. Without such provenance, models may be executable but difficult to verify or trust.

Diagnostic checking and correction should follow an explicit inspect-check-repair-verify loop. MCP-enabled tools first inspect generated objects and simulation logs. Schema, geometry, envelope, system and physical plausibility check then identify missing references, invalid geometry, implausible properties, control inconsistencies, severe warnings or numerical instability. The LLM may classify the issue and propose a bounded correction, but a model-editing tool applies the change and records the affected object, evidence and rationale. Validation and simulation tools then rerun the relevant checks. The loop ends when predefined acceptance criteria are met or a modeller needs to review an uncertain or consequential change. In this loop, the LLM plans the sequence, MCP standardises how each tool is discovered and invoked, and the invoked tools record the resulting actions; none of these replaces diagnostic rules or professional judgement. This approach aligns with the need for physical consistency and interpretability in data-driven

building simulation.¹⁰

CONCLUSION

MCP should not be framed as a shortcut around BEM expertise. Its promise lies in making BEM workflows more connected, inspectable and reproducible. Recent work on LLM-based BEM, multi-agent retrofit modelling, EnergyPlus-MCP, MCP-enabled agentic workflows and RAG-MCP urban simulation shows that the field is moving from isolated demonstrations towards protocol-mediated modelling infrastructure.

A trustworthy agentic BEM workflow should begin with the anatomy of the model, engage with heterogeneous building information, support bounded information-to-BEM generation, validate and repair model objects, and expose each assumption as part of the modelling record. In this framework, the language model is neither author nor oracle. It is a bounded coordinator of tools, data and judgement.

Future studies should develop a shared MCP-enabled BEM ecosystem and registry comprising data connectors, weather and material services, BEM engine adapters, validation routines, calibration tools and visualisation components. Standardised tool schemas, units, provenance fields, permissions and versioning would allow resources from different developers to be combined and audited. Common benchmark tasks should report accuracy, failure cases, human time, runtime and reproducibility materials, including tool schemas, model and weather files, assumptions, outputs, validation scripts and action logs. Such infrastructure would support interoperability while preserving explicit validation, traceability and professional accountability.

If developed with explicit validation, rigorous provenance and controlled autonomy, MCP may help move BEM beyond isolated expert file manipulation towards more connected, reproducible and scalable modelling practice for low-carbon building design and operation.

REFERENCES

1. Liu M., Zhang L., Chen J., et al. (2025). Large language models for building energy applications: Opportunities and challenges. *Build. Simul.* **18**:225–234. DOI:[10.1007/s12273-025-1235-9](https://doi.org/10.1007/s12273-025-1235-9)
2. Arslan M. and Munawar S. (2026). Large language models in building energy applications: a survey. *Energy Build.* **352**:116800. DOI:[10.1016/j.enbuild.2025.116800](https://doi.org/10.1016/j.enbuild.2025.116800)
3. Jiang G., Ma Z., Zhang L., et al. (2024). EPlus-LLM: A large language model-based computing platform for automated building energy modeling. *Appl. Energy* **367**:123431. DOI:[10.1016/j.apenergy.2024.123431](https://doi.org/10.1016/j.apenergy.2024.123431)
4. Zhang L., Ford V., Chen Z., et al. (2025). Automatic building energy model development and debugging using large language models agentic workflow. *Energy Build.* **327**:115116. DOI:[10.1016/j.enbuild.2024.115116](https://doi.org/10.1016/j.enbuild.2024.115116)
5. Li H., Zhang L., Zhou H., et al. (2026). MCP-enabled agentic AI workflow for building energy modelling: framework and use cases. *J. Build. Perform. Simul.*:1–27. DOI:[10.1080/19401493.2026.2653969](https://doi.org/10.1080/19401493.2026.2653969)
6. Li H., Xu Y. and Hong T. (2025). EnergyPlus-MCP: A model-context-protocol server for ai-driven building energy modeling. *SoftwareX* **32**:102367. DOI:[10.1016/j.softx.2025.102367](https://doi.org/10.1016/j.softx.2025.102367)
7. Ciccozzi A., Santavica A., de Rubeis T., et al. (2025). BIM-BEM interoperability for energy analysis: A comparative study of different strategies. *Energy Rep.* **13**:4705–4718. DOI:[10.1016/j.egy.2025.04.027](https://doi.org/10.1016/j.egy.2025.04.027)
8. Lu J., Zheng Z., Langtry M., et al. (2025). Automated building energy modeling for energy retrofits using a large language model-based multi-agent framework. *iScience* **28**:113867. DOI:[10.1016/j.isci.2025.113867](https://doi.org/10.1016/j.isci.2025.113867)
9. Zhou J., Chen J., Wang M., et al. (2026). Large language model orchestrated workflow integrating retrieval-augmented generation and model context protocol tools for urban-scale simulation-based resilience assessment. *Build. Environ.* **297**:114553. DOI:[10.1016/j.buildenv.2026.114553](https://doi.org/10.1016/j.buildenv.2026.114553)
10. Jiang Z., Wang X., Li H., et al. (2025). Physics-informed machine learning for building performance simulation – A review of a nascent field. *Adv. Appl. Energy* **18**:100223. DOI:[10.1016/j.adapen.2025.100223](https://doi.org/10.1016/j.adapen.2025.100223)

DECLARATION OF INTERESTS

The authors declare no competing interests.