Article

Not peer-reviewed version

The Glass Box Planner: From Deductive Policy Audits To a Normative Framework for Urban AI

Kiarash Mohammadi *

Posted Date: 1 June 2026

doi: 10.20944/preprints202606.0063.v1

Keywords: land use compatibility; computational planning; policy-as-code; deductive audit; algorithmic governance; explainable AI (XAI)

Article

The Glass Box Planner: From Deductive Policy Audits To a Normative Framework for Urban AI

Kiarash Mohammadi

Independent Researcher, Iran; kiarash7.mohammadi@gmail.com

Abstract

This paper makes two contributions. First, it bridges the land use analysis gap by replacing manual methods with a scalable, open-source engine implementing a transparent 'policy-as-code' approach. We applied the Compatibility Audit Tool (CAT) to Qazvin, Iran, analyzing over 65,000 land parcels and revealing that a critical 2.04% of the urban fabric—concentrated at residential-industrial interfaces—was in direct policy conflict. The framework provides planners with a robust instrument for a systematic 'policy audit' to identify contradictions between policy and reality. Second, it proposes a normative framework for urban AI, shifting from optimization-focused models toward forensic instruments that enforce accountability by quantifying the divergence between stated policy and spatial reality. It transforms the planning audit from a bureaucratic formality into a mechanism for liability discovery.

Keywords: land use compatibility; computational planning; policy-as-code; deductive audit; algorithmic governance; explainable AI (XAI)

---

1. Introduction

1.2. The Challenge of Land Use Compatibility

Land use conflicts in modern cities challenge sustainable growth (Berke, 2006). While planning support systems incorporating GIS (Yang et al., 2008) and remote sensing enable spatial analysis and modeling, their limitations have driven a movement to codify planning regulations into computable "Rules as Code" (Barry and Daniel, 2021). This approach translates legislative logic into computable code (Mowbray et al., 2023). Our work applies this principle to urban planning, creating a transparent procedural instrument for systematic evaluation (Stead, 2021) which we term the ***Compatibility Audit Tool (CAT)***. Our work aligns with 'Governance by Glass-Box' (Aler Tubella et al., 2019), but with a distinction: our tool is an intrinsically transparent engine for deductive auditing, not an external monitor for an opaque system.

Yet, this trajectory toward transparency is complicated by the rise of urban AI, where "participation" is often framed superficially. As a recent review by Sieber et al. (2025) demonstrates, the discourse is dominated by models that treat civic input as a mere byproduct of automation, a problem of quantification, or a technocratic exercise in building trust. This paper presents a methodological and philosophical counterpoint, proposing the ***Glass Box*** as a normative framework to guide the development of such tools.

1.3. From Value Overlays to Inductive Simulation: A Counter-History of PSS

Although the "Rules as Code" concept is very new (Barry and Daniel, 2021), its reasoning is similar to the deductive transparency of McHarg's landscape overlays from 1969. However, this early readability was subordinated to analytical complexity in the later development of Planning Support Systems (PSS), as demonstrated by land-use transport interaction (LUTI) models (Yang et al., 2008), cellular automata (White and Engelen, 1997), and Agent-Based Models (ABM) (Batty et al., 2012). This trajectory widened the "implementation gap" noted by Geertman (2008), separating advanced technologies from real-world utility. By going back to transparent, logical reasoning and eschewing predictive simulation in favor of a methodical audit of the city's stated aims, the Compatibility Audit Tool (CAT) addresses this gap.

1.4. Two Modes of Urban AI: Instrumental Tools vs. Normative Actors

We must distinguish two applications of urban AI; failing to do so fuels the techno-solutionist fallacy permeating smart city discussions.

AI as an Instrumental Tool performs targeted, practical tasks whose results can be compared to an objective ground truth. The fundamental question is: 'How can we most efficiently achieve this predefined goal?' Examples include classifying land cover from sensor data or using computer vision to detect building footprints from satellite imagery.

The more dangerous mode, AI as a ***Normative Actor***, influences ethical, moral, and social decisions. Rather than merely learning facts about the world, the AI is projecting about how it ought to be, asking "What is the best way for us to live together?" An AI that "optimizes" a city budget or proposes a new zoning plan is one example of this. This mode cannot be compared to any objective ground truth. Democratic contestation and debate are the only credible checks.

Our central argument aligns with Chantal Mouffe’s (2005) **critique of the "post-political" consensus.** We contend that abusing instrumental tools to "optimize" the city is an attempt to stifle legitimate conflict. While Habermas (1984) feared the "colonization of the lifeworld" by technical systems, Mouffe warns that masking disagreement behind technocratic consensus causes latent conflict to explode into **antagonism**.

This distinction is at the heart of our 'Glass Box' framework. By making conflict clear and controllable—turning the "enemy" into an "adversary"—the tool is designed to stop this slide into antagonism. The Compatibility Audit Tool (CAT) embodies this principle: its embedded norms—the compatibility matrix—are perfectly explicit, legible, and therefore contestable. It is an instrumental tool built specifically to structure and serve the normative, democratic process, ***providing the forensic evidence necessary to hold that process accountable.***

1.5. A Spectrum of Intervention: Distinguishing AI in Design, Planning, and Governance

The discourse over 'smart cities' and 'urban AI' is often dangerously imprecise, conflating distinct modes of computational intervention, each with its own scale, function, and political risks.

AI in Urban Design:

At the building or site scale, AI's function is that of a **generative assistant**. At the architectural level, AI can enable creativity by rapidly generating and evaluating thousands of formal options based on quantifiable metrics. However, the final normative judgment of ‘good’, ‘beautiful’, and ‘fitting’ remains with the expert and the community. In this domain, the AI operates on geometry and form, optimizing for defined parameters without infringing on the political rights of the inhabitant.

AI in Urban Planning:

At the city-wide scale, a new paradigm of LLM-driven planning has emerged, often proposing to simulate or automate the inherently political process of participation itself (Ni et al., 2025; Zhou et al., 2024). By attempting to solve the "city-as-lived" *in silico*, these models risk hallucinating consent. Our work presents a deliberate counter-proposal: the AI as a **consequence engine**. Its legitimate role is not to make new plans or simulate citizens, but to enable **forensic audits** of the systemic, spatial consequences of existing, human-written policies. This ensures the tool serves to audit the *policy* rather than replace the *political subject*, making the political landscape more transparent and accountable. This is the domain our methodology is designed to occupy.

AI in Urban Governance:

AI at this scale acts as an automated administrator, managing operational decisions like policing or traffic. This is the most dangerous application, carrying high risks of systemic bias, surveillance, and erosion of democratic accountability (Taeihagh, 2021). It supports a depoliticization of urban

 doi:10.20944/preprints202606.0063.v1

governance by collapsing contentious political choices into supposedly neutral, instantaneous calculations (Olmstead, 2025), which launders systemic bias into code. This is a direct instance of the 'colonization of the lifeworld'.

We must distinguish between these roles to prevent importing the high-risk, 'black box' paradigm of governance into the necessarily deliberative and political work of planning. Recent scholarship proposes a typology that moves from AI-assisted roles to fully AI-automated or even AI-autonomized planning (Peng et al., 2024). Our goal is to move beyond a technical typology to a ***political one***, rejecting a linear path toward automation that even its supporters acknowledge must struggle with planning's basis in human "values, judgments, and preferences".

1.6. Beyond the Digital Twin: Auditing the 'Low-Frequency City'

'Digital twin' has gained prominence, but Batty (2018) argues the concept is ambiguous, often promising a logically impossible city 'mirror image'. He draws a critical distinction between the ***'high-frequency city'*** of real-time operations and the ***'low-frequency city'*** where the long-term, strategic work of planning unfolds. The current paradigm, exemplified by the "Glass Box Adaptation" for Bologna's digital twin (Costagliola et al., 2025), focuses on the former: creating complex, multilayered simulation models to manage high-frequency urban networks like transportation and energy.

Our tool is not a real-time digital twin but a purpose-built model for the ***'low-frequency city'***. Its function is not to mirror a city's dynamics, but to audit the slow, accretive, and often contradictory outcomes of its foundational planning policies. This approach avoids the conceptual paradoxes of the digital twin while providing a robust instrument for the core task of urban planning.

1.7. Operationalizing Compatibility: The Matrix as a Political Artifact

To operationalize compatibility, the CAT uses a matrix to analyze land use relationships. The matrix codifies a community's planning goals, cultural values, and legal precedents. In short, the matrix is an ***inherently political artifact***, translating normative goals into computable rules.

This subjectivity is not a flaw; it is the engine of **Agonistic Pluralism**. August (2024) critiques agonistic theory for failing to provide concrete "practices of conflict regulation" or institutions that can handle deep disagreement. **The Compatibility Matrix answers this deficit.** By forcing vague values into explicit code, the Compatibility Matrix transforms a nebulous "clash of civilizations" into a tangible, line-by-line negotiation. It serves as the **institutional vessel** that Mouffe (2005) demands, providing the precise mechanism needed to sublimate raw conflict into structured political debate.

This forces a direct confrontation with what Cornwall (2008) identifies as the central problem of participation: its status as an "infinitely malleable concept". By demanding this codification, the CAT resists the "vagueness" that allows participation to become a tool of legitimation rather than genuine empowerment. The tool's legitimacy rests on the robustness of the democratic process used to create it (Arnstein, 1969), ensuring the framework is human-legible, its assumptions are contestable, and its logic is derived from the situated expertise of local professionals.

For a policy to be equitable, it must be applied consistently, yet the sheer scale of modern cities makes manual audits based on a compatibility matrix impractical, error-prone, and subject to inconsistent enforcement based on institutional memory or political pressure.

1.8. The Inductive Scout vs. The Deductive Audit

To bridge this operational gap, our **CAT** expedites the analysis based on the curated matrix. Using ML algorithms risks laundering spatial injustices into seemingly objective models. For example, models trained on market data simply amplify the logic of capital—it would code for gentrification, privileging profitable adjacencies, and penalizing uses that don't generate immediate economic return, regardless of their social value (Gordon, 2019).

This temptation toward automated judgment leads to a dominant, yet philosophically distinct, workflow: the use of AI as a *statistical scout*. This is exemplified by approaches that use complex machine learning models and rely on post-hoc interpretation methods to explain the "black box" after the fact (Papadakis et al., 2024). Such reliance on post-hoc interpretation methods is a barrier to the responsible application of AI in the profession. The inherent opacity of inductive models creates a "black box" that planners cannot question, audit, or trust (Sanchez, 2025).

These models are designed to analyze numerous metrics—such as property values, business revenues, or noise complaints—to identify areas of statistical interest for expert review. However, this approach's limitation is not its analytical power but its potential for misinterpretation. Such models highlight areas based on statistical correlations and outliers found within the data—an inductive, bottom-up approach to *discover* conflict from emergent patterns (Brown and Raymond, 2014). This acts as agenda-setting, directing experts to focus on problems visible and quantifiable in chosen datasets, while potentially overlooking significant issues that are not captured by those metrics. While valuable for exploratory analysis, this inductive workflow becomes problematic when its data-driven findings are mistaken for impartial policy judgments. The inductive scout operates on the lower rungs of Arnstein's (1969) 'ladder of citizen participation.' At best, it offers a form of 'consultation' or 'informing' by highlighting statistical curiosities for experts. It does not, however, equip communities with the power to contest the underlying policy itself.

Our '**Glass Box**' framework is built on a **deductive** foundation. We term this a ***Deductive Audit.*** It resolves a core tension in the '**Rules as Code**' field: the choice between representing what rules are *interpreted to do* versus what they *literally say* (Mowbray et al., 2023). Rather than an inductive model discovering emergent patterns, the CAT is a deductive engine for policy auditing. It asks not ***'Where are the interesting patterns?'*** but rather, ***'Where does the on-the-ground reality contradict the explicit propositions of our stated plan?'*** This reframing creates a tool to ascend Arnstein's ladder, moving beyond tokenism by providing auditable evidence for genuine debate. The transparency is neither post-hoc nor algorithmic; it is foundational and human-legible from the start. For the high-stakes, regulatory context of urban planning—a domain that Papadakis et al. (2024) note is increasingly governed by mandates like the EU AI Act—we argue this ***"transparent by design"*** approach, which emphasizes logical legibility over the behavioral compliance monitoring central to other 'Glass-Box' models (Aler Tubella et al., 2019), is more robust, legitimate, and directly aligned with the principles of democratic accountability.

The compatibility matrix explicitly codifies a community's planning goals, cultural values, and legal precedents. The CAT does not find 'interesting' patterns; it **systematically identifies contradictions** between a community's stated intentions—itself a product of communicative action—and its on-the-ground reality. It respects the boundary between the two rationalities: the **CAT** is a systems-based tool designed not to replace but to structure and inform the democratic debate of the lifeworld. This provides a more robust and legitimate starting point for professional inquiry, focusing an expert's judgment on resolving *known policy conflicts* rather than on interpreting *ambiguous statistical patterns*.

1.9. A Normative Framework for Future Urban AI

Our critique is not a rejection of complex multi-input AI. The potential of a deep neural network, one capable of analyzing multiple, disparate data sources—from the built environment, the natural environment, the human environment, and the "ground truth" for conflict is undeniable. However, moving toward such an integration demands that we squarely face three foundational issues:

1.10. The Data Challenge

What data are we feeding it? What data are we feeding it as a proxy to unquantifiable data? Are we violating the privacy and data sovereignty of the public? These models excel at pattern-seeking and recognition, but this means that high-risk data like raw mobility data (GPS traces), emergency service calls, and individual crime reports must never enter the main model in its raw state. Even census data that is considered generally public must be carefully handled on a granular level, such as where a small number of individuals live on a block and the model would be explicitly getting that

 doi:10.20944/preprints202606.0063.v1

data. Separate ethical review boards must be established and citizens must have a right to agree or disagree to data use agreements for such systems that are moving beyond the traditional analytics.

1.11. The Normative Challenge

How and for which community are we optimizing it? By what values, and in whose image are we making it? In a purely technical lens, an AI optimizing for public commute routes might cut a single stop with low foot traffic to improve efficiency. What if that single stop is exactly what a limited number of elderly people and children use for their commute? Are we taking the AI at face value or are we investigating the underlying humanistic aspect? The choice of objective function is inherently political: *Maximizing Tax Revenue*, *Minimizing Infrastructure Costs*, *Minimizing Service Load*, or *Minimizing Political Friction* each have their own uses and pitfalls. In the absence of a single correct image to build an AI on we must answer: whose image are we building the AI in?

Perhaps the answer must be *'**every image**'*, implying the need for multiple AIs. The comparison of their results would become the very grounds for a more educated democratic discourse and informed decision-making.

1.12. The Epistemological Challenge

Our critique of singular, optimization-focused models is not merely a caution for the future; it is a lesson from the past. As Batty (2024) reflects, early computational planning's quest for optimization was largely abandoned precisely due to its inability to account for critical aspects of urban life that can't be easily quantified. This includes concepts like ***social cohesion***—the density of trust, mutual support, and social bonds in a neighborhood. It includes ***perceived safety***—not the statistical likelihood of being a victim, but the feeling of being unsafe which dictates behavior and degrades quality of life. It includes the subjective beauty, charm, or *"vibe"* of a place. And it includes ***political capital***—the ability of a community to advocate for its interests and resist unwanted change. While proxies can be used to gesture at these qualities—such as voter turnout, 'Eyes on the Street' indices, or demographic velocity—a more valuable resource is direct engagement through community surveys that serve as a Human-in-the-Loop element that might highlight the actual feeling of the community and contrast with the data.

While no single tool can solve these foundational issues, our ***Glass Box framework*** provides a scaffolding for addressing them. It directly confronts the ***Normative Challenge*** by requiring that values be made explicit and contestable within the compatibility matrix. It provides a powerful starting point for the ***Epistemological Challenge*** by generating a map of spatial hypotheses that can guide targeted, qualitative research into unquantifiable realities. Finally, it establishes a guardrail against the ***Data Challenge*** by operating on a foundation of parsimonious, public data, setting a standard of data minimalism before more complex and high-risk datasets are ever considered.

2. Methodology

The CAT is a high-performance, vectorized Python engine utilizing GeoPandas. The pipeline requires a geospatial dataset and a compatibility matrix. We utilized the 2012 Qazvin parcel dataset, which remains the most recent comprehensive, open-access spatial record available for the city. In contexts of high institutional opacity, where real-time data is often nonexistent or classified, the ability to perform retrospective audits on available 'static' data is critical. Therefore, this dataset serves as a robust demonstration artifact for the methodology, proving that the CAT can extract actionable policy insights even in verified, albeit non-contemporary, data environments. Our findings represent a methodological proof-of-concept, not a current diagnostic. Data was standardized to the matrix taxonomy and re-projected to the local UTM.

2.1. Constructing the Provisional Matrix

While the matrix should ideally be co-created through deliberative democracy (Healey, 1997), we constructed a provisional matrix as a ***demonstration artifact*** based on four heuristics derived from Qazvin's master plan documents (see Supplementary Material A):

- *Protecting Sensitive Uses:* Health and residential zones were coded as incompatible ("1") with high-externality uses like industrial;
- *Self-Compatibility:* All diagonal relationships were coded as fully compatible ("5");
- *Default Neutrality:* Undefined relationships defaulted to neutral ("3"); and
- *Fostering Synergies:* Vibrant adjacencies, such as residential near green space, were coded as fully compatible.

2.2. The Computational Process

The engine employs a vectorized workflow. Unique parcel IDs are assigned, and adjacency is determined via Euclidean buffers. We selected a 10m buffer to operationalize immediate adjacency (shared boundaries/alleys) and a 25m buffer to represent the immediate neighborhood context.

2.3. Equity & Data Scarcity

While network-based distances or view-sheds offer greater precision, they demand sophisticated, often proprietary datasets (e.g., complete topological road networks) frequently unavailable in the contexts this tool is designed to serve. We selected Euclidean buffers to ensure accessibility in data-scarce environments, forcing a transparent decision about analysis scale rather than obscuring it behind complex dependencies.

A spatial self-join identifies all neighboring parcel pairs. Finally, a 'worst-case scenario' aggregation assigns the minimum compatibility score to any parcel facing multiple adjacencies. This tuning for high sensitivity ensures no potential policy violation goes unnoticed.

2.4. Outputs and Inherent Limitations

The tool generates a GeoPackage for visualization and summary CSV reports detailing compatibility performance across land use categories.

Crucially, the CAT audits the statutory city (current or draft plan) rather than the city-as-lived. It serves as a "representation of space" (Lefebvre, 1992), detecting syntactic inconsistencies between the city and its regulations. It cannot resolve semantic ambiguity or the "open texture" of urban life (Mowbray et al., 2023). Thus, a "policy conflict" identified by the audit indicates a logical contradiction to be investigated, not necessarily a failure of urban quality. It is a starting point for inquiry, not a final judgment.

2.5. Implementation and Open Science

The analysis engine and a standalone GUI are implemented in Python. To ensure reproducibility, the annotated source code, matrix, and datasets are available in: https://github.com/Kiarash-m0hammadi/glass-box-planner.

3. Results

This map shows that while most of the city exhibits high compatibility, areas of conflict are concentrated in the southwest corner of the city, with some along a major arterial corridor in the center and in the northwest part of the development from the last two decades.

 doi:10.20944/preprints202606.0063.v1

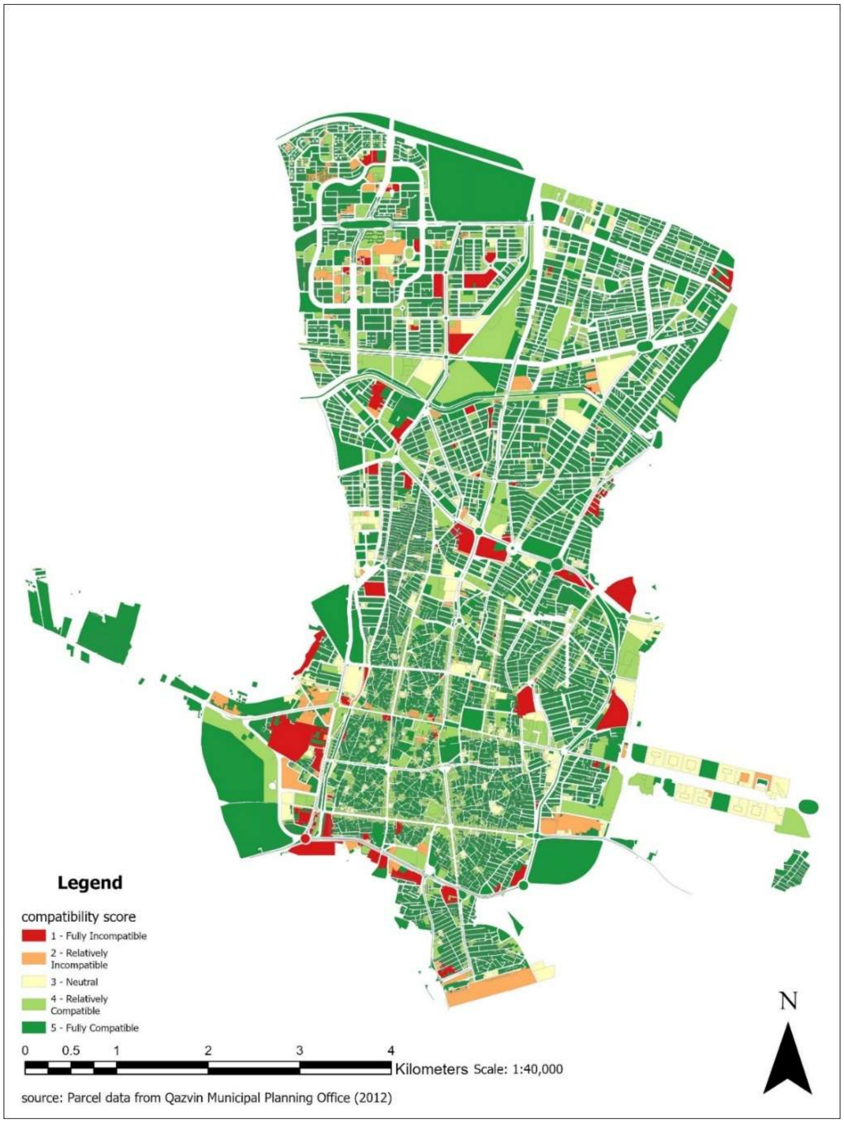

Figure 1. Spatial Distribution of Land Use Compatibility in Qazvin, Iran.

As detailed in Table 1, while only 2.04% of the urban fabric (1,329 parcels) is classified as incompatible, this includes 697 residential parcels and 28 schools. This small percentage represents a significant volume of acute conflict, disproportionately affecting sensitive uses.

Table 1. Compatibility analysis summary at 10 and 25 meters.

Compatibility Score	10m Buffer		25m Buffer	
	Parcel Count	Percentage of Total Parcels (%)	Parcel Count	Percentage of Total Parcels (%)
1 - Fully Incompatible	748	1.15	1,444	2.22
2 - Relatively Incompatible	581	0.89	1,139	1.75
3 - Neutral	1,350	2.07	2,874	4.42
4 - Relatively Compatible	7,738	11.89	13,129	20.18
5 - Fully Compatible	54,652	83.99	46,483	71.44
Total	65,069	100.00	65,069	100.00

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2026 doi:10.20944/preprints202606.0063.v1

This cartographical form of conflict is not representative of the informal "street ballet" (Jacobs, 1961). An incompatible area might be a thriving community, while a formally compatible area might suffer from issues. This disconnect was highlighted by local research in Qazvin, which found a gap between the neighborhood's formal layout and residents' actual 'sense of security' (Nazari et al., 2015) has highlighted might be suffering from issues. This distinction between ***auditing the plan*** and ***understanding the place*** is paramount.

3.1. Hotspot Analysis: From District-Level to Parcel-Level

Figure 2 illustrates the tool's capacity to identify specific typologies of conflict. Figure 2a isolates residential parcels lacking enforced buffering against Urban Utilities, while Figure 2b exposes a "sandwich" condition where residential blocks abut a legacy workshop zone. By identifying the exact parcels subject to negative externalities (e.g., noise, heavy logistics), the CAT moves beyond district-level generalizations to pinpoint precise locations for mitigation or rezoning.

Figure 2. Acute Land Use Conflict between Residential and Urban Utilities and Infrastructure Parcels in the central area.

Figure 3. A Hotspot of Severe Incompatibility Caused by the Adjacency of Residential Housing and an Industrial Workshop Zone.

 doi:10.20944/preprints202606.0063.v1

3.2. Sensitivity Analysis: The Critical Role of Scale

As shown in **Table 1**, expanding the buffer to 25m increases the number of incompatible parcels from 1,329 to 2,583, nearly doubling the identified conflicts. This highlights how critical the definition of "neighborhood scale" is to the outcome of the audit.

4. Discussion

4.1. The Matrix as a Forensic Instrument: From Debate to Accountability

The matrix codifies the dynamic 'place-shaping continuum' (Carmona, 2014), making the outcomes of political power relationships ***explicit and auditable.***

A "Shadow Audit" is made possible by this legibility. The official zoning matrix (Matrix A) and an external standard, like WHO guidelines (Matrix B), can run concurrently. The difference becomes a measurable indicator of regulatory carelessness, not just a difference of opinion. This enables advocates to inquire as to why the local "law" deviates from the standard of care, taking the tool beyond simple compliance monitoring and transforming it into a diagnostic for institutional failure.

The political dimension necessitates a more thorough, open, and democratic process for creating the matrix itself. To organize stakeholder discussion, the matrix should use established MCDM methods (Soltani et al., 2015). In particular, a method such as the AHP offers a reliable and methodical way to extract and measure participants' subjective, pairwise assessments, turning their values into a common ledger (Schmoldt et al., 1995). This transforms the black box of the 'expert-defined' matrix into a shared ledger of community values. Communities and NGOs could easily propose their own counter-matrices based on their own values, generating their own cartographic argument. The two levels this audit operates are against the stated intentions of the plan. Second, and more profoundly, it audits the plan itself questioning the consequences of a potentially outdated, "petrified" matrix of rules.

Beyond a single, generalized matrix, each neighborhood could have its own matrix for instance, prioritizes heritage preservation or social infrastructure over commercial intensity as a form of bottom-up approach framing disagreements.

The CAT aims not to manufacture consensus, but to map Agonism. Following Mouffe (2005), democracy seeks not to eliminate conflict, but to establish 'conflictual consensus' on the rules. The CAT provides those rules. By visualizing exactly where the 'City as Planned' fights with the 'City as Lived,' the tool does not solve the dispute, but it ensures the dispute is fought with **forensic evidence** rather than populist rhetoric.

The abstract codified matrix shows the on-the-ground spatial consequences for all to see. Examine the hotspot shown in Figure 2b, which consists of a zone of industrial workshops sandwiched between a strip of residential blocks. The Glass Box map makes it impossible to overlook the effects of previous zoning decisions, but it does not resolve this issue on its own. It presents a particular, fact-based argument to Qazvin's residents and planners: Should the city spend a lot of money on mitigation and buffering? Should it develop an industrial relocation long-term plan, and if so, who pays for it? Or should it think about rezoning the residential lots to reflect the district's industrial nature? The **CAT** uses a straightforward, cartographic argument to raise these precise political issues.

Sieber et al. (2025) identified a gap in civic AI: the authors' 'reluctance to explicitly include political power.' Our method addresses this by ***shifting the burden of proof.*** In a traditional public meeting, a resident's claim that a neighborhood is 'unsafe' is treated as subjective anecdote. By quantifying the gap between the plan's promises and the ground truth, the **CAT** converts that anecdote into **forensic evidence of policy violation.** It transforms the planning environment from an abstract clash of values to a litigious reality where the institution must justify its failure to adhere to its own code.

 doi:10.20944/preprints202606.0063.v1

4.2. Bridging the Plan and the Place: The Audit as a Starting Point

The **CAT's** limitation is also its greatest strength: it audits the ***city as planned***, or what Lefebvre (1992) called "representations of space," rather than the city as lived. This distinction is methodologically significant. Just as an identified "policy conflict" does not necessarily indicate a dysfunctional neighborhood, a "high compatibility" score does not equate to a high quality of life. As a result, the tool's output is not an absolute opinion; It serves as a springboard for research and a mapping tool intended to close the gap between the two.

A map of spatially explicit hypotheses is the **CAT's** main output. The "green zones" of formal compatibility and the "red zones" of formal conflict are places to start a more thorough, qualitative examination of the lived experience; they are the most critical locations to begin deeper investigation. This creates a powerful mixed-methods workflow:

Phase 1: Deductive Audit. The Glass Box identifies where the on-the-ground reality contradicts (or conforms to) the explicit logic of the city plan.

Phase 2: Targeted Qualitative Inquiry. Planners and researchers can then use this map as a sampling frame for targeted fieldwork—ethnographies, go-along interviews, or community workshops—in these specific zones. The research questions become precise and place-based:

- **In a 'green zone'**: "The plan dictates that this neighborhood should be harmonious. Does it feel that way? Does it, despite its formal compatibility, suffer from the low 'sense of security' identified in Qazvin by Nazari et al. (2015)?"
- **In a 'red zone'**: "The plan suggests this area is in direct conflict. Why does it (or does it not) function? What informal practices, social networks, or Jane Jacobs' (1961) 'street ballet' are mitigating the formal policy contradictions?"

The "Epistemological Challenge" that was previously mentioned is directly addressed by this workflow. Unquantifiable elements, such as "social cohesion" and "perceived safety," are not attempted to be measured by the **CAT**. Rather, it employs the plan's logical structure to pinpoint the exact places where those lived realities—which cannot be measured—have the greatest impact. It offers a useful instrument for organizing and guiding the essential task of qualitative, human-centered planning research. It tells us exactly ***where to go to ask the right questions, not what the answer is.***

4.3. Resisting the Leviathan: The Glass Box as an Alternative to Algorithmic Governance

The Glass Box framework addresses a critical sequencing issue in AI governance: moral boundaries cannot be imposed through "Governance by Glass-Box" (Tubella, et al., 2019) if they were not first established through a contestable democratic process. The demise of Sidewalk Labs offers definitive proof of this error. Public criticism accurately noted that corporate-led, opaque policies infringed on data sovereignty (Wang, 2021), framing residents merely as "consumers" at the lowest rungs of participation (Cardullo and Kitchin, 2019).

Our framework reverses this logic. Rather than managing residents through opaque efficiency metrics, the CAT provides them with auditable evidence—derived from the plan's own logic—to contest the city's trajectory. It asserts that tools must function to support political agency, not replace it.

However, recent proposals for "synthetic participation" require a precise distinction. Frameworks that create closed-loop "Planning, Living, and Judging" cycles between AI agents (Ni et al., 2025; Zhou et al., 2024) commit a significant category error. By substituting a frictionless, *in-silico* pantomime for the messy legitimacy of real-world communicative action, these methods attempt to simulate away the Normative Challenge.

Yet, a total rejection of simulation risks overlooking the structural biases of human participation itself. As Einstein et al. (2019) demonstrate in *Neighborhood Defenders*, the traditional public meeting is often dominated by a distinct minority—older, wealthier homeowners—while renters and the working class remain structurally absent. Therefore, we propose a role for synthetic agents not as a **substitute** for the citizen, but as a **"Red Team" for the consensus**. By simulating the

 doi:10.20944/preprints202606.0063.v1

perspectives of the absent to stress-test the decisions of the present, the AI ensures that the Glass Box facilitates a debate that is truly inclusive, preventing the process from becoming a vehicle for either algorithmic *or* demographic bias laundering (Zheng et al., 2025).

4.4. Future Work: Towards a Hybrid, Deliberative AI

Distinguishing the CAT's deductive audit from inductive discovery is not an argument for rejection, but for ***necessary sequencing.*** The deductive audit is the essential first step—the establishment of a legible social contract—within which inductive tools can then be safely explored.

Future work could integrate deep learning to capture complex, non-linear interactions between noise, traffic, and green infrastructure—relationships deterministic models like Mansourihanis et al. (2023) struggle to represent. A proposed solution is coupling it with a 'GIS-native' Explainable AI (XAI) framework. This suggests a powerful synthesis of the two workflows. Such a system would not only highlight areas of potential conflict using the deductive 'Glass Box' method but would also allow for a subsequent inductive analysis to generate a series of disaggregated factor maps, allowing planners to interrogate the model's reasoning and understand precisely which factors are driving the conflict in a specific location.

LLMs have immense potential if deployed as instrumental tools serving, rather than replacing, human communicative action. Potential use cases might include: **Translator** for jargon, a **Brainstorming Partner** for design options, a **Synthesizer** for workshop analysis, and an **Agonistic Red-Teamer** to challenge consensus by representing marginalized voices.

An LLM configured to represent specific marginalized demographics can analyze proposed plans or meeting transcripts to flag 'silent' grievances. This prevents the sanitization of conflict and ensures that the 'unheard' perspective is entered into the forensic record, even if those citizens could not physically attend.

In these modes, the AI acts as cognitive scaffolding for human debate—translating and organizing thought without usurping the normative authority to define values.

This hybrid approach allows for consideration of real-world consequences while holding the plan accountable. This would represent a paradigm shift from **diagnostic to prescriptive analytics**, creating a powerful synergy between the predictive power of AI and the deliberative needs of democratic planning.

This discourses favors moving away from black boxes and towards an era of explainable AI (XAI). This entails creating a pipeline where pure algorithmic tools work in series, producing not a single answer, but a series of snapshots interpretable by professionals and the stakeholders alike.

4.5. Generalizability and Conclusion

The **CAT's** architecture—a user-defined pairwise matrix coupled with a high-performance vectorized spatial analysis—is a broadly generalizable framework. beyond conventional land use planning, a variety of urban phenomena can be easily analyzed thanks to its architecture. To model retail synergies, for instance, the land use categories could be swapped out for particular business types, and the matrix would account for competitive friction or similar customer demographics. In urban design, it could be used to evaluate how well various building masses or architectural styles blend together in a historic district. The **CAT** offers a fundamental methodology for a novel class of open, expert-led spatial analyses.

The proposed tool is not to achieve frictionless, optimized urbanism, but as a diagnostic tool for highlighting conflict, laying the groundwork for an evidence-based democratic argument about the city's future. This paper has therefore advanced a dual contribution: the ***Compatibility Audit Tool (CAT),*** a pragmatic, open-source tool for deductive policy auditing, and the ***Glass Box,*** a conceptual framework for computational planning. This deductive approach is offered in contrast to two other dominant modes: the use of participatory methods like PPGIS **purely for inductive conflict discovery**—that is, to identify emergent spatial patterns of community-perceived conflict (Brown & Raymond, 2014)—and the complex simulations of technocratic models that quantify conflict between predefined ecological, production, and living functions (Zou, et al., 2021). Its goal is to supplement professional judgment by exposing the real-world effects of explicit, human-defined policies—not to replace it with an algorithm.

According to the dominant perspective, planners can concentrate on more complex tasks since AI has "freed them from drudge work" (Sanchez, 2025). This viewpoint Ignorantly overlooks the main way planners acquire tacit knowledge and a thorough comprehension of the "city-as-lived" is through the so-called "drudge work"—the laborious, manual process of auditing plans and interacting with the finer details of the city. Our framework re-centers this work as the central epistemological engine of a reflective and democratically accountable planning practice, rather than as a tedious task to be automated away, by offering a tool for a more methodical and transparent audit.

However, turning policy into code risks algorithmic petrification: freezing consensus in time to create rigid digital bureaucracy (Winner, 2017). A plan on paper can be changed; a rule in code is a brutal ossification of logic. The **CAT** must therefore be integrated into the continuous civic process encouraging debate about the matrix itself; using it as a static end point amplifies the danger of petrification.

Properly applied, the tool functions as a **'zoning debugger'** for active planning. Just as a software compiler checks code for syntax errors before execution, the CAT allows a planner to audit a **draft master plan** against the compatibility matrix. This preemptive forensic step identifies 'silent errors'—such as an accidental industrial buffer violation hidden in a complex new district—**before the plan is legislated.** This allows the planner to correct mistakes in the 'drafting phase' rather than fixing them in the 'built phase,' effectively serving as a spell-checker for spatial policy.

This transforms the **CAT** from a purely diagnostic instrument into a dynamic ***decision-support tool*** for evaluating policy alternatives. This transformation establishes the foundational layer of democratic legitimacy, making the Glass Box not just a tool, but a prerequisite for the ethical use of any subsequent AI in planning. Ultimately, however, for managing conflicts, a whole other toolset including buffers, landscaping buffers, or noise mitigation walls, aesthetic overhauls and direct capital investment must be implemented that closes the gap between computation, policy, and the city on the ground.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Land Use Compatibility Matrix.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aler Tubella A, Theodorou A, Dignum F, et al. (2019) Governance by Glass-Box: Implementing Transparent Moral Bounds for AI Behavior. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, California, August 2019, pp. 5787–5793. International Joint Conferences on Artificial Intelligence Organization.
2. Arnstein SR (1969) A Ladder Of Citizen Participation. *Journal of the American Institute of Planners* 35(4): 216–224.
3. August V (2024) Understanding democratic conflicts: The failures of agonistic theory. *European Journal of Political Theory* 23(2): 182–203.
4. Barry R and Daniel C (2021) Introduction to Rules as Code in Planning. Available at: https://www.linkedin.com/pulse/introduction-rules-code-planning-richard-barry/.
5. Batty M (2018) Digital twins. *Environment and Planning B: Urban Analytics and City Science* 45(5): 817–820.
6. Batty M (2024) AI and design. *Environment and Planning B: Urban Analytics and City Science* 51(4): 799–802.
7. Batty M, Crooks AT, See LM, et al. (2012) Perspectives on Agent-Based Models and Geographical Systems. In: *Agent-Based Models of Geographical Systems*. Dordrecht: Springer Netherlands, pp. 1–15.
8. Berke P (2006) *Urban Land Use Planning*. University of Illinois Press. Available at: https://books.google.com/books?id=bSlUAAAAMAAJ.

 doi:10.20944/preprints202606.0063.v1

9. Brown G and Raymond CM (2014) Methods for identifying land use conflict potential using participatory mapping. *Landscape and Urban Planning* 122: 196–208.
10. Cardullo P and Kitchin R (2019) Being a 'citizen' in the smart city: up and down the scaffold of smart citizen participation in Dublin, Ireland. *GeoJournal* 84(1): 1–13.
11. Carmona M (2014) The Place-shaping Continuum: A Theory of Urban Design Process. *Journal of Urban Design* 19(1): 2–36.
12. Cornwall A (2008) Unpacking 'Participation' Models, meanings and practices. *Community Development Journal* 43(3): 269–283.
13. Costagliola A, Montanari M and Bellavista P (2025) Interconnecting Urban Networks: A Novel Approach to Digital Twins Through GlassBox Adaptation. In: *Proceedings of the 14th International Conference on Smart Cities and Green ICT Systems*, 2025, pp. 193–204. SCITEPRESS - Science and Technology Publications.
14. Einstein KL, Glick DM and Palmer M (2019) *Neighborhood Defenders: Participatory Politics and America's Housing Crisis*. Cambridge University Press. Available at: https://books.google.com/books?id=UYHCDwAAQBAJ.
15. Geertman S (2008) Planning Support Systems (PSS) - a planner's perspective. *Planning Support Systems for Cities and Regions*. Epub ahead of print November 2008.
16. Gordon F (2019) Virginia Eubanks (2018) Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: Picador, St Martin's Press. *Law, Technology and Humans*: 162–164.
17. Habermas J and McCarthy T (1984) *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*. Beacon Press. Available at: https://books.google.com/books?id=16ntg728ZdAC.
18. Healey P (1997) *Collaborative Planning: Shaping Places in Fragmented Societies.* s.l.:Macmillan.
19. Jacobs J (1961) *The Death and Life of Great American Cities*. New York: Random House.
20. Lefebvre H (1992) *The Production of Space*. Wiley. Available at: https://books.google.com/books?id=SIXcnIoa4MwC.
21. Mansourihanis O, Maghsoodi Tilaki MJ, Yousefian S, et al. (2023) A Computational Geospatial Approach to Assessing Land-Use Compatibility in Urban Planning. *Land* 12(11): 2083.
22. McHarg IL (1969) *Design with Nature*. Garden City, NY: Natural History Press.
23. Mouffe C (2005) *On the Political*. Routledge.
24. Mowbray A, Chung P and Greenleaf G (2023) Representing legislative Rules as Code: Reducing the problems of 'scaling up'. *Computer Law & Security Review* 48: 105772.
25. Nazari V, Keshavarz S and Omidi P (2015) Urban space adaptation to reduce social anomalies and increase the sense of security (Case study: Hadi Abad neighborhood, Qazvin)[مناسب سازی فضای شهری به منظور کاهش ناهنجاری های اجتماعی و افزایش احساس امنیت (نمونه موردی: محله هادی آباد قزوین)]. In: *First Specialized Conference on Architecture and Urban Planning of Iran*, Shiraz, Iran, 2015. Civilica. Available at: https://civilica.com/doc/411198.
26. Ni H, Wang Y and Liu H (2025) *Planning, Living and Judging: A Multi-agent LLM-based Framework for Cyclical Urban Planning*. Available at: www.aaai.org.
27. Olmstead NA (2025) Sidewalk Toronto and the discursive politics of the real-time city. *Time & Society*. Epub ahead of print 22 April 2025. DOI: 10.1177/0961463X251333857.
28. Papadakis T, Christou IT, Ipektsidis C, et al. (2024) Explainable and transparent artificial intelligence for public policymaking. *Data & Policy* 6: e10.
29. Peng Z-R, Lu K-F, Liu Y, et al. (2024) The Pathway of Urban Planning AI: From Planning Support to Plan-Making. *Journal of Planning Education and Research* 44(4): 2263–2279.
30. Sanchez T (2025) Will Urban Planners Lose Their Jobs to AI? Available at: https://www.planetizen.com/blogs/136043-will-urban-planners-lose-their-jobs-ai.
31. Schmoldt D, Peterson D and Smith R (1995) The analytic hierarchy process and participatory decisionmaking. *Proceedings of the 4th International Symposium on Advanced Technology in Natural Resources Management*. Epub ahead of print November 1995.
32. Sieber R, Brandusescu A, Sangiambut S, et al. (2025) What is civic participation in artificial intelligence? *Environment and Planning B: Urban Analytics and City Science* 52(6). SAGE Publications Ltd: 1388–1406.

doi:10.20944/preprints202606.0063.v1

33. Soltani A, Hewage K, Reza B, et al. (2015) Multiple stakeholders in multi-criteria decision-making in the context of Municipal Solid Waste Management: A review. *Waste Management* 35: 318–328.
34. Stead D (2021) Conceptualizing the Policy Tools of Spatial Planning. *Journal of Planning Literature* 36(3): 297–311.
35. Taeihagh A (2021) Governance of artificial intelligence. *Policy and Society* 40(2): 137–157.
36. Wang J (2021) Why Did Sidewalk Labs's Quayside Project Fail? The Practical and Institutional Difficulties of Smart City Practice. In: November 2021.
37. White R and Engelen G (1997) Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B: Planning and Design* 24(2): 235–246.
38. Winner L (2017) Do artifacts have politics? In: *Computer Ethics*. Routledge, pp. 177–192.
39. Yang F, Zeng G, Du C, et al. (2008) Spatial analyzing system for urban land-use management based on GIS and multi-criteria assessment modeling. *Progress in Natural Science* 18(10): 1279–1284.
40. Zheng Y, Xu F, Lin Y, et al. (2025) Urban planning in the era of large language models. *Nature Computational Science* 5(9). Springer Nature: 727–736.
41. Zou L, Liu Y, Wang J, et al. (2021) An analysis of land use conflict potentials based on ecological-production-living function in the southeast coastal area of China. *Ecological Indicators* 122: 107297.
42. Zhou Z, Lin Y, Jin D, et al. (2024) Large Language Model for Participatory Urban Planning. Epub ahead of print 27 February 2024.