Beyond Machine Perception: AI Urban Imagination

Darío Negueruela del Castillo, Iacopo Neri

Introduction

The city has always been a subject of fascination and study, its complexity and dynamism defying simple categorization. In recent years, the rapid development of AI has opened up new avenues for exploring the urban landscape, offering novel tools and the capacity to analyze multimodal data at scale, which could contribute to our understanding of cities as a complex, historically grounded, and culturally mediated environment.

This paper delves into the potential of AI to go beyond mere machine perception and engage with the city as a "found world," shaped by both material structures and contested fields of vision, imagination, and representation. It examines three projects: Clip and the City (2023), World Gist (2023–24), and The City in a Bottle (2024–25), each leveraging different AI techniques to analyze urban landscapes, uncover cultural biases, and simulate urban environments through the lens of critical urban theory. These systems, in their very attempt to render the city legible through algorithmic means, paradoxically perform what Baudrillard would recognize as a parade of simulacra, where the model precedes and determines the territory it is supposed to represent.

The Closure of Urban Imagination

Our age seems dominated by "platform urbanism."[1] The city becomes increasingly indistinguishable from its digital shadow, cast by the vast apparatus of data collection and algorithmic processing that mediates urban experience. This apparatus extends its reach into economic, leisure, social, and affective spheres, shaping the dynamics of branding, retail, property development, and services that give form to our urban environments. This already widens the gap between lived experience and comprehension of the larger structures that shape that experience, but there is more to the story.

AI foundation models and multimodal Large Language Models (LLMs) now join this picture. These generalist models, trained on a huge snapshot of the Internet, like Common Crawl[2] (itself a kind of digital unconscious of contemporary culture), are increasingly present, mediating our online and physical lives. They perform a restriction of urban imagination to the combinatorial possibilities inherent in their training distributions, which I call "algorithmic foreclosure of the possible." This operates through a recursive semiotic trap. The AI systems, in learning from the sediment of Internet culture, construct what appears to be an infinite space of potential but which is, in reality, a closed system, a hermetically sealed universe of remixed existing forms. These models' latent spaces have been called accidental archives or snapshots of our culture, but are in fact self-referential learned distributions with no outside space, where the "culture" that is encoded is trapped within its own walls. Moreover, these latent spaces have become the main brokers of our visual imaginaries, conditioning the generation of synthetic visuals that flood the Internet. This represents not merely a limitation of current technology but perhaps also a fundamental contradiction within the logic of computational urbanism itself.

---

1   Platform urbanism describes the intersection of digital platforms and urban environments. It represents an evolution of the Smart City paradigm, focusing on how digital technologies and services reshape urban spaces and social interactions in which large tech platforms represent a new kind of political geography, and specifically, a new kind of sovereignty, one that potentially transcends the local and national.

2   See https://commoncrawl.org.

Surplus Data and Urban Knowing

Previously, the paradigm of Big Data, where large amounts of data coming from our digital footprints, smartphones, and from the ubiquity of sensors, promised to deliver the structure behind all kinds of phenomena by letting data speak and avoiding imposed or preconceived theories. This went hand in hand with the ideas behind the Smart City model.[3] But we are now facing a different reality. We seem to have too much data, not too little. In what we could consider the third phase of "machinic urbanization" (following Mumford's paleotechnic and neotechnic orders), we are now in a new paradigm that Orit Halpern has called "surplus data." This emerges not merely as an accumulation of information but as a fundamentally new mode of understanding and therefore acting upon cities, a new urban epistemology. Halpern has observed that "data is reassembled and recombined to produce new 'truths' from within the system. Data extends beyond description, creating the world it would describe."[4] In other words, new and autonomous worlds emerge from this vast amount of available and ingested data, especially in large AI models. This takes on a rather sinister aspect for architecture and urbanism, as data becomes a "perfect crime" (Baudrillard) against urban reality itself, eliminating not just the referent but the very possibility of reference.

It is useful here to describe a triad of projects that try to tackle the hidden world of architecture within AI foundation models. With these projects, we attempt to analyze what implicit ideas these models have of architecture. Or in other words: what do they think of architecture, and what are their implicit models and theories?
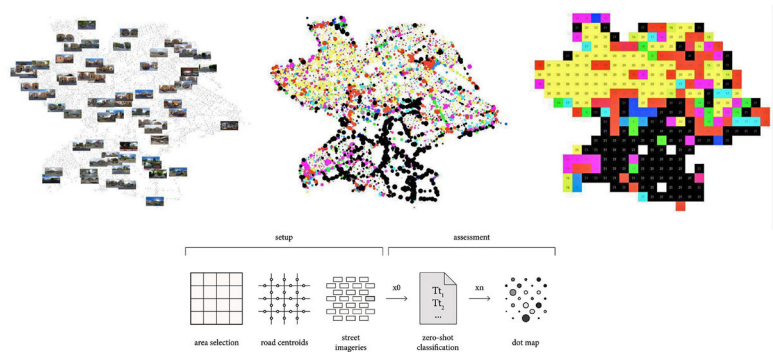
Clip and the City: Uncovering Urban Cultural Layers

The probabilistic logic of AI systems exemplified in projects like CLIP and the City[5] represents not simply a new tool for urban analysis but a transforma-

---

3   For an informed view, see Rob Kitchin, "The Real-Time City? Big Data andSmart Urbanism," *GeoJournal* 79, no. 1 (2014): 1–14, https://doi.org/10.1007/s10708-013-9516-8.

4   Orit Halpern and Robert Mitchell, *The Smartness Mandate* (MIT Press, 2022), 203.

5   Darío Negueruela del Castillo, Iacopo Neri "CLiP and The City: Addressing the Artificial Encoding of Cities in Multimodal Foundation Deep Learning Models," in *On Architecture* (2023) Conference Proceedings (Strand, 2023), 92–99, https://doi.org/10.60152/zkhvd3 og.

tion in what constitutes urban knowledge itself. In this project, we fed 360° urban panoramas of Rome[6] into large vision-language models (CLiP and Open-CLIP), to extract the embeddings[7] of their visual and textual features. This enabled us to perform spatial analyses that revealed clusters and discontinuities in the urban layout. These clusters, in turn, highlighted the interplay of historical legacies and contemporary urban practices. The project ultimately underscored that while machine learning models can capture remarkable details of urban form, they also propagate cultural biases inherited from their training datasets.

Fig 27: Diagram showing the pipeline for the project, first we download google street imagery for the full city centre of Rome, then we query the matching of these images with labels from urban and architectural terminology through the model CLiP, finally, we spatialize back the result in a map of the city to assess.

As we have demonstrated through our computational pipeline, these systems operate within the cultural logic of computational urbanism, where the city becomes less a physical reality than a set of statistical correlations floating in high-dimensional space. The physical reality, its dimensions of encounter, embodied knowledge, different registers of self and other, are subordinated to that abstract high-dimensional embedding space. Our findings prompt a more
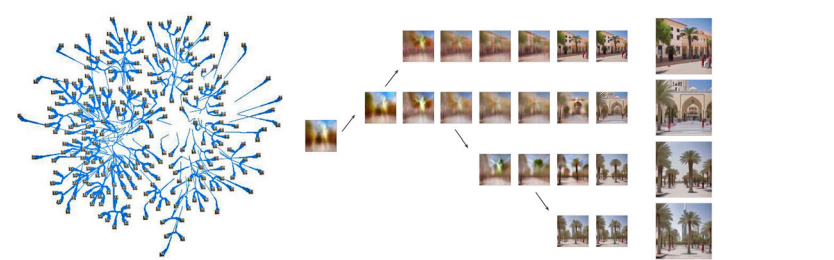
---

6   The city of Rome was chosen as our primary case study due to its rich historical tapestry and cultural resonance in Western thought.

7   Which give us information about their relative position with respect to other images and text, revealing proximities and similarities that uncover the way in which the models have "learnt" about the city of Rome.

cautious integration of such models in urban studies, advocating for a foren-
sic reading of the city where historical narratives and machinic interpretations
intersect.

World Gist: Mapping the Global Urban Imagination

*Fig. 28: Genealogy of the generation in WorldGist, showing the different paths in the
denoising process, leading to different cities*

Shifting our focus from a single city to a global perspective, the World Gist
project[8] reverses the scale logic of our earlier experiment. Here, we harnessed
the power of Stable Diffusion conditioned by CLIP to generate synthetic im-
agery for every capital city worldwide. By keeping the same seed—that is, the
same noise distribution—we ensured that all image generations were subject
to an identical starting point. This methodological choice allowed us to isolate
how the model's latent urban knowledge varies across different cultural and
geographical contexts. World Gist reveals a kind of computational unconscious
of contemporary urbanism. Our experiments revealed that the model "knows"
more about cities that are emblematic of globalized culture—predominantly
those in the Global North or cities that have achieved a high degree of cultural
saliency. These clusters suggest that the model organizes urban representa-
tions around certain generic centers of gravity in its latent space. The insights
from World Gist expose the uneven distribution of urban knowledge within
AI models, revealing what Jameson would call the "political unconscious" of

---

8   Dario Negueruela del Castillo and I. Neri, "World Gist" [research paper examining how
     Stable Diffusion and CLIP encode and reproduce urban knowledge] (Routledge, forth-
     coming).

global urbanization, the repressed contradictions and uneven development that haunt our planetary urban system.[9]

The City in a Bottle: Simulating Urban Assemblages

While both previous projects focused on the representation of cities from existing imagery or synthetic generation, The City in a Bottle takes a bold step toward simulation. This project bridges the gap between data-driven digital twinning and critical urban theory by leveraging advanced vision-language models enhanced with Retrieval-Augmented Generation (RAG). Rather than simply replicating urban phenomena through aggregated data, our method simulates the multifaceted assemblages that constitute contemporary cities, integrating a discursive reasoning on social, material, and physical context. We operationalized seminal urban theories from authors like Henri Lefebvre, Jane Jacobs, Walter Benjamin, and Georg Simmel, as well as from contemporary frameworks such as assemblage theory and urban metabolism. For instance, we extracted key concepts from Jacobs' insights like "urban choreography of play" and distilled them into perceivable visual features that could act as measurable proxies for testing theories on urban imagery. This involved extracting a concept description in cases where it was missing, devising which questions we could ask to evaluate the presence in an image (an urban panorama) of urban elements that make it possible, and breaking it down into a series of reasoning steps that render it graspable and measurable (see appendix). The hybrid method allows us to "bottle" the cognitive processes of urban theorists alongside the tangible urban fabric, offering a speculative yet grounded tool for urban interpretation. Initial results demonstrate that by blending computational precision with theoretical insight, we can generate simulations that highlight dynamic interactions among human and non-human actors, the evolving nature of urban resilience, and the possibilities for climate adaptation.

---

9   Frederic Jameson, *The Political Unconscious: Narrative as a Socially Symbolic Act* (Cornell University Press, 1981).

Fig. 29: Emulating a narrative urban flânerie through the city of Madrid in the think-ing style of Henri Lefebvre from the project The City in a Bottle.

The Triple Abstraction in AI's understanding of the City

This condition of surplus data cannot be understood apart from what we might call the cultural logic of late computational capitalism, following Jameson. The very accumulation of urban data represents a new form of primitive accumulation, transforming lived urban experience into abstract, commodifiable data points. This process, far from being neutral or purely technical, embodies a type of "data colonialism,"[10] a new phase in the long history of extractive relationships to urban space and experience.
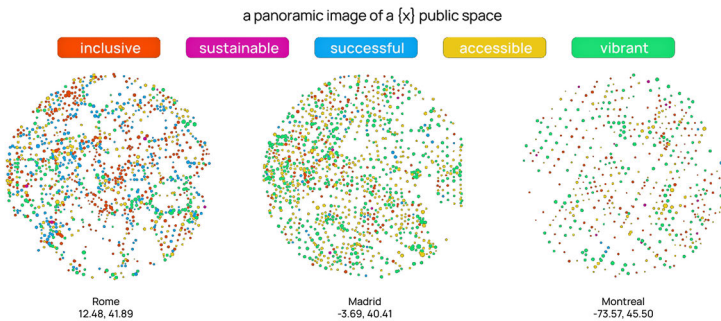
The three projects under consideration—CLIP and the City, World Gist, and The City in a Bottle—are not simply applications of AI to urban problems, but reshape the very nature of urban knowledge. The probabilistic logic they employ, while appearing to embrace uncertainty and multiplicity, paradoxically performs a closure of the urban imagination, which is also a digital enclosure of the urban commons of the mind. What occurs in these times of algorithmic urbanity is not merely a quantitative expansion of urban data but a qualitative transformation in the very possibility of urban imagination itself.

---

10   Nick Couldry and Ulises A. Mejias, "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject," *Television & New Media* 20, no. 4 (2018): 336–49, https://doi.org/10.1177/1527476418796632.

Here we confront the fundamental contradiction of computational urbanism:[11] the simultaneous hypermapping and erasure of urban possibility. The more thoroughly our AI systems map and model urban space, the more completely they chase away the genuinely new.

This foreclosure operates through a triple movement of abstraction. First, the reduction of lived urban experience to data points (what Mumford would recognize as the ultimate victory of abstraction over the organic and concrete); second, the transformation of this data into high-dimensional latent spaces that, while mathematically infinite, remain topologically closed; and finally, the generation of new urban "possibilities" that are, in fact, merely probabilistic recombinations of existing forms. The CLIP and the City project reveals this process with particular clarity. We performed a zero-shot classification of urban scenes, and while this appears to uncover new urban meanings, it actually performs algorithmic reification: it transforms living urban relations into fixed, computable categories. The project's findings reveal not just bias but that perfect crime against urban reality we mentioned earlier: a substitution of the map for the territory so complete that the territory almost vanishes entirely.

Fig. 30: The methodology of the CLIP and the City project allows us to compare the machinic imaginary of cities at scale using available urban panoramas and conduct comparative analyses. In the image: visualizing the scores of urban terminology in Rome, Madrid, Montreal at the sacle of the city centre.

a panoramic image of a {x} public space

inclusive   sustainable   successful   accessible   vibrant

Rome
12.48, 41.89

Madrid
-3.69, 40.41
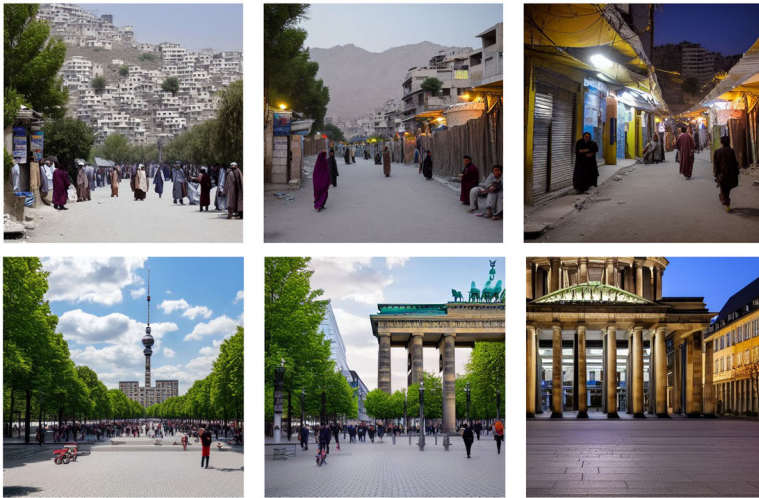
Montreal
-73.57, 45.50

---

11   This extends Jameson's analysis of postmodernity: Frederic Jameson, *Postmodernism, or, the cultural logic of late capitalism* (Duke University Press, 1991).

The Paradox of AI's Urban Knowledge

World Gist's attempt to construct global urban imaginaries through machine learning confronts us with a kind of a paradox. The more comprehensively these systems process urban data, the more thoroughly they eliminate the "regional framework of civilization" (to quote Mumford again); that is to say, the specific, the local, the unreproducible. This is not merely a technical limitation but rather the logical culmination of the "age of the world picture,"[12] in the words of Martin Heidegger, where the world becomes available to us only as a computed representation.

Fig. 31: WorldGist: Generating images of cities during the day, evening and night, starting from the same seed (same noise distribution as point of departure)

Yet perhaps it is precisely in this moment of apparent total closure that the possibility of genuine urban imagination might be found again. We can, for instance, follow Jameson's insistence on the political unconscious, and identify in the very contradictions of algorithmic urbanism the seeds of what could go

---

12   Martin Heidegger, "The Age of the World Picture," in *The Question Concerning Technology and Other Essays*, trans. William Lovitt (Harper and Row, 1977), 115–54.

beyond. The City in a Bottle project, in its attempt to simulate historical urban thinkers' perspectives on contemporary cities, inadvertently reveals the limits of computational urban thinking, the "blind spots of the algorithm."

What is required is not an abandonment of computational tools but rather a "right to the algorithmic city," extending Lefebvre's ideas of the "right to the city"[13] and putting forward a demand that urban AI systems be repurposed toward what Walter D. Mignolo terms "digital pluriversality."[14] This would entail not merely technical modifications but a fundamental rethinking of what constitutes urban knowledge itself. Such a project would need to embrace productive contradictions: the tension between data's promise of total knowledge and the irreducible particularity of urban experience; between algorithmic prediction and genuine urban emergence; between the closure of computational thinking and the openness required for genuine urban imagination.

New Urban Possibilities

This brings us finally to the political dimension of urban AI, the "distribution of the urban sensible."[15] The question is not simply how to make AI systems more accurate or comprehensive, but how to preserve and nurture spaces of genuine urban imagination within and against the totalizing logic of computational urbanism. This perhaps requires going beyond the divide between human understanding and machine prediction and embracing an unavoidable hybrid condition. We could call this, extending Haraway's concept, an "urban cyborg politics," one that neither fully embraces nor entirely rejects algorith-

---

13   A similar idea has been put forward by Rashid Mushkani, Hugo Berard, Allison Cohen, and Shin Koeski, "Position: The Right to AI," ArXiv, January 29, 2025, https://doi.org/10.48550/arXiv.2501.17899.

14   Walter D. Mignolo, *The darker side of Western modernity: Global futures, decolonial options* (Duke University Press, 2011).

15   Jacques Rancière, *The Politics of Aesthetics: The Distribution of the Sensible*, trans. Gabriel Rockhill (Continuum, 2004). For a discussion and reinterpretation of Rancière's ideas for our times of algorithmic governmentality, see Antoinette Rouvroy and Bernard Stiegler, "Le régime de veritéé numérique," Socio. La nouvelle revue des sciences sociales, no. 4 (2015): 113–40.

mic mediation, but rather seeks to redirect it toward genuinely emancipatory ends.[16]

In this context, AI foundation models like CLIP manifest what we have identified as the technological unconscious of contemporary urbanism. These systems, in their attempt to parse and classify urban imagery, reveal not just the visible and salient aspects of cities but also the accumulated visual ideology of late capitalism encoded in their training data.[17] The zero-shot classifications they produce are thus less discoveries than symptoms, revealing the implicit urban theories embedded in the Internet's visual culture.

The City in a Bottle project presents perhaps the most telling contradiction: In its attempt to simulate how canonical urban thinkers might perceive contemporary cities, it reveals what Baudrillard would call the "precession of simulacra" in urban theory itself.[18] The project's use of RAG systems to ventriloquize historical urban theorists performs double simulation, of both urban space and of urban thought itself. What becomes clear through this analysis is the need for a dialectical urban AI.[19] This would entail, first, the recognition of the importance of the fragments of urban reality that like computational ruins resist algorithmic processing; second, the development of urban AI models that can behave more like "organic prediction machines," modes of computational analysis that enhance rather than suppress urban complexity; and, third, the cultivation of counter-algorithmic practices as methods of urban imagination that explicitly work against predictive closure.

Conclusion: Beyond the Digital Closure of Urban Imagination

What emerges from this analysis is not merely a critique of computational urbanism but what we might term, following Jameson, a fundamental contradiction in the very project of algorithmic urban knowing. Here, the crucial aspect is not simply the revealing of unnoticed patterns due to enhanced machinic

---

16   Donna Haraway, "A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century," in *Simians, cyborgs, and women: The reinvention of nature* (Routledge, 1991), 149–81.

17   Jameson would recognize this as their repressed cognitive content.

18   Jean Baudrillard, *Simulacra and Simulation*, trans. Sheila Faria Glaser (University of Michigan Press, 1994).

19   Following Benjamin's concept of dialectical images: see Walter Benjamin, *The Arcades Project*, trans. Howard Eiland and Kevin McLaughlin (Belknap Press, 1999).

perception, but rather the alignment of the reasoning capacities of the models with the interpretative and narrative threads proposed by scholars of reference, and which are already a work of analytical and critical synthesis. The condition of surplus data, which promised to make cities more legible, predictable, and manageable, is paradoxically producing a crisis of urban imagination. This crisis manifests not as a scarcity of possibilities but as their algorithmic foreclosure, a situation where, to paraphrase Mark Fisher, it becomes easier to imagine the end of cities than to imagine genuine urban alternatives outside the predictive logic of AI systems.[20]

The three projects examined here—CLIP and the City, World Gist, and The City in a Bottle—reveal different aspects of this crisis. As we have shown in our study of the latent urban imaginaries embedded within AI models, these systems are not merely tools but active participants in shaping how cities are imagined, designed, and developed. They reveal a blurry generative urban grammar that determines which versions of urban futures can be articulated, thereby performing an enclosure of imagination.

Yet perhaps it is precisely in these limitations that the possibility of genuine urban imagination might be rediscovered, in what we might call the computational ruins of contemporary urbanism. This would require what we could term a "dialectical urban AI," one that neither fully embraces nor entirely rejects computational mediation, but rather works through its contradictions toward new forms of urban possibility. The challenge, then, is not simply to critique these systems but to repurpose them toward a new "right to the algorithmic city," extending Lefebvre's ideas of the right to the city.[21] This can be conceived as the right not just to access urban data but to participate in the very definition of what constitutes urban possibility, from the categories and labels used to categorize data, the architecture of the algorithms used to extract features, all the way to the way latent spaces are articulated into insights, intellectual accompaniment, or production. This would entail the recognition of what exceeds computational capture—which aspects of urban life resist algorithmic reduction—the development of counter-practices that work against the grain of predictive closure, and the cultivation of a more organic and serendipitous urban computation as ways of using AI that enhance rather than suppress the vital and emergent qualities of urban life.

---

20   Mark Fisher, *Capitalist Realism* (Zero Books, 2009).

21   A similar idea has been put forward by Mushkani et al., "Position: The Right to AI."

As we navigate this moment of technological transformation, the question becomes not simply how to make our AI systems more accurate or comprehensive, but how to preserve and nurture spaces of genuine urban imagination within and against the totalizing logic of computational urbanism. For it is only through such a dialectical engagement with our technological condition that we might move beyond the digital closure of urban imagination.