Synthetic Data Meets Architectural Typology: An Exploratory Computational Workflow with a Carbon Footprint Inference Case Study

DANIEL KOEHLER
University of Texas at Austin, School of Architecture

Keywords: Generative AI, building typology, compositional analysis, synthetic data, environmental design.

This study employs large-scale-text-to-image (LLI) models to analyze building typologies in relation to environmental contexts. Here, types delineate deep compositions of places that intertwine socio-economic histories with physical structures. Rather than mere representations or models, types encapsulate various facets of a site into a specific term, acting as strategic interfaces bridging architectural design and policy-making. Moving forward, can typological thinking assist in understanding generative-AI workflows from an architectural perspective? Moreover, can one redesign types as instrumental interfaces once again linking design to their environmental contexts?

The investigation examines the compositional characteristics of AI-generated images of buildings across various cities. A synthetic dataset of 150,000 images was segmented into individual building segments, enabling a statistical analysis of compositional features across 5,600 cities. The paper introduces how LLI models portray diverse local typologies and differentiates these using computational metrics for big-data analysis. It explores the LLI’s potential to assess the carbon footprint of places by analyzing materials, building parts, and construction methods within the generated images through image segmentation.

Despite only a real-world alignment of approximately 70 percent in the synthetic data, such databases can augment existing building datasets. Synthetic datasets are particularly useful in hard-to-access contexts and in past or projective settings. They allow for the precise staging of specific content, such as building typologies or perspectives. Computational articulation of types reveals specific attributes that transcend linear classification regimes, aiding here in assessing places’ carbon footprints through multilayered linkages. The analysis indicates that embodied carbon in places does not align with geographical carbon classifications, offering more differentiated resolutions. Furthermore, the embodied carbon, composed of multiple materials, is not visually apparent, suggesting a need for local and project-specific adaptations.

INTRODUCTION

Generative design is a methodology that automates the creation of design options, balancing a variety of competing goals. The latest iteration is driven by Large Language models (LLMs), better known as Generative AI. Unlike traditional AI systems that primarily define decision boundaries, Generative AI emphasizes artifact production and displays versatility in generating multiple valid outputs for a given input. Traditional machine learning mainly centers on data classification according to particular criteria. However, embracing the concept of Generative Variability, Weisz et al envision design as an expedition emphasizing multiple outputs and their subsequent analysis. Design becomes an exploration and the challenge to navigate through a latent space spanned by trillions of tokens of data. A notion very familiar to architects accustomed to orchestrating the construction of buildings from literally millions of individual bricks. This study also conceives generative design as an investigative procedure, accentuating compositional attributes over fixed objectives. The typological examinations are data-centered, emphasizing data collection, augmentation, and comparison to discern patterns.

This study applies large-scale-text-to-image (LLI) models to investigate building typologies, focusing on their relationship with socioeconomic and environmental contexts. The findings indicate that Generative AI, especially when using synthetic data, bears a profound resonance with Aldo Rossi’s seminal theories on architectural type. Rossi’s influential work conceptualized the city as a continual construction from the geography of its architecture, underscoring the concept of permanence. Just as Rossi saw urban artifacts and building types as permanent, unchanging fixtures that order the city, generative models trained on diverse urban data create computational types that encode urban knowledge. Synthetic training data allows generative networks to learn the visual language of urban elements across eras, from classical facades to modernist housing blocks to postmodern interiors. As models internalize the motifs and rhythms of building parts, plazas, and streetscapes, they gain an implicit understanding of the urban environment as a geography of persisting, identifiable types arrayed in unique but intelligible combinations. This computational typology could help steer...
urban adaptation along more coherent, humane lines aligned with Rossi’s vision of the city as a spatial work of architecture.

Rossi and his contemporaries called the New Rationalists leveraged the knowledge on type as the basis for its re-composition. They recognized the transformative potential of reconfiguring parts and wholes within diverse configurations and settings. Building elements are intricately entwined with their contexts, encompassing historical indices or material cultures, which frequently present virtual and stark contrasts to their locations. Consequently, the compositional elements of a type intersect with high-level features of buildings such as spatial organization, functional program, construction techniques, and socio-political milieu. Rossi’s interpretation of city architecture hinged on leveraging enduring elements to craft strategic junctures, thereby bridging architectural design and policy-making. Analogously, generative models capitalize on amassed urban insights to suggest future adaptations of an urban motif. Through Generative AI, one could revive the theoretical framework of the recomposition of a building, its context, and its parts by understanding their local permanence and the implications of inserted exceptions.

STATE OF THE ART

Recent literature on generative AI’s impact on architecture has painted a predominantly pessimistic portrait, emphasizing concerns about the potential obsolescence of architects’ and even declaring the “death of the architect.” This paper advocates for a more optimistic yet realistic perspective on AI’s capabilities, highlighting its fundamental divergence from previous technologies adopted in the architectural field. Much critique relies on a limited conception of the architect as a mere draftsperson. Yet historically, notations enabled architects to delegate construction and focus on qualitative aspects of the built environment. In this vein, large language models (LLMs) emerge as innovative notational tools, freeing architects to channel more resources toward enhancing qualitative facets. The transformative capacity of generative AI presents not an existential threat, but an opportunity to elevate the architect’s purpose by reducing time spent on drawing. While acknowledging valid concerns, I emphasize AI’s potential to build on, not supplant, human creativity and critical thinking. Further interdisciplinary research is needed to delineate appropriate applications of generative models that augment, rather than replace, the architect’s unique capacities. Overall, a nuanced examination of the continuities and ruptures between past technical shifts and current AI reveals significant unrealized potential.

Large language models are exhibiting increasing versatility as they rapidly assimilate new expertise through appropriate data training and prompting. For instance, designers now employ ChatGPT to generate operational plugins, suggesting potential
integration of software development into the design process and reduced reliance on rigid workflows. Architectural drawing becomes a more fluid endeavor, able to be individually tuned by personal AI assistants. Rather than a dystopian vision of large language models as monolithic, corporation-controlled super-architects, these AIs emerge more as interfaces. They facilitate swift domain expertise access, translation, and customization based on individual requisites. In sum, large language models’ adaptable enhancement, not replacement, of human creativity reimagines architectural practice. The fluidity enabled by AI systems allows architects to operate with enhanced responsiveness to qualitative inputs, social needs, and environmental factors. This refocuses the architect’s purpose on human-centric value creation over the production of drawings as an end itself.

Recent research indicates language models can infer physical interactions and spatial representations, blurring boundaries between language and physicality. For instance, language models can understand and predict expected motions and physical relationships between entities, whether a violin falling or the movement anticipated in a conversation. This comprehension extends to spatial domains like architecture. Given a scene description, OpenAI’s GPT4 can generate a corresponding plan, exhibiting spatial understanding and potential for image generation that bridges text and physicality. For example, the fine-tuned version “September 25” of ChatGPT-4 can analyze uploaded plan drawings and describe alignment with ADA regulations, even suggesting design adjustments.

In summary, generative models comprehend diverse informational languages, enabling translation across knowledge domains. Focusing on the interaction of knowledge types, this study wants to explore whether new insights on architectural typology can emerge from generated data. The fluidity of mapping information across modalities, from language to spatiality, presents new potentials for typological investigation.

**METHODS**

A critical precedent for this research is the work of Areti Markopoulou and Oana Taut, who propose reading the city as a circular material resource, facilitated through machine vision and the creation of a database of existing building stock. Building on this, the study segments a generated database of approximately 150,000 images of residential buildings across 5,600 cities worldwide each with over 100,000 inhabitants to draw insights on material usage and their embodied carbon. The research involves several steps: first, compiling a list of appropriate cities and building typologies; second, generating a dataset of building images; third, segmenting each image into labeled object masks; fourth, aggregating segment features into data sheets; fifth, formulating compositional measures to visualize the data; and finally, conducting a case study to derive insights on environmental factors, such as carbon footprint, from prevalent features of a city’s building types.

In the initial research phase, a list of cities was compiled from the open-source World Cities Database by Simplemaps, which sources data from authoritative agencies like NGIA, USGS, Census Bureau, and NASA. The selection criteria included every city with a population over 100,000. The focus was on generating images of buildings that represented prevalent housing and residential structures in each location. For this purpose, a specific prompt was generated for each city following the structure “a large (...) in”, with (...) being replaced by the most common building terms at that location. This was achieved by mapping their probabilities for each specific city using the OpenAI GPT2 API. Such a method stimulated more context-specific building arrangements in a continuous composition. Furthermore, prompts were confined in terms of perspective, lighting, and photorealism to isolate factors of formal composition.

For image generation, the Stability AI’s Stable Diffusion 2.1 model was utilized due to its open-source nature, which allows for script customization. Its neutrality and lack of stylistic bias, in contrast to platforms like Midjourney, made it a preferable choice. A stochastic search process was employed to identify 40 optimal seeds that accurately represented locale-specific
building types while maintaining a balance between variabil-
ity and cross-city comparability. Subsequent outlier removal
reduced the sample to 24 images per city. Outliers included
instances of multiple images in one and non-side-view perspec-
tives. It was observed that identical seeds with varied city names
could yield significantly different images, reflecting context sen-
sitivity. However, further constraints like fixed outlines using
tools such as ControlNet were deliberately avoided to maintain
the variability of the generated images.

The synthetic image dataset was segmented using the Meta
Segment Anything and Groundino models, generating approxi-
mately 150 distinct masks per image encoded in RLE format.
This format is a machine-readable structured text that stores
metadata in addition to geometric descriptions. Masks de-
lineate the shape of specific objects within the image, enabling
the analysis of individual segments. The study focused on iden-
tifying crucial features for architectural analysis by segmenting
the images into four categories: quantities, geometric features,
materials, and building parts.

The Segment Anything model initially segments images se-
mantically, differentiating objects in a scene. This enables the
collection of quantitative data like segment size, count, or area,
and geometric features such as eccentricity, which measures a
segment’s roundness. These data points are combined to derive
representative compositional features. For example, segment
count estimates the number of building parts present, while
corner count, obtained by polygon approximation, indicates
architectural complexity. Average segment area reveals typical
part sizes across cities, while the area standard deviation shows
the variety of sizes. Higher values indicate greater variability.
Additionally, average compactness, quantified as the area-to-
perimeter ratio squared, assesses overall shape complexity
across segments. Simpler shapes like circles have higher values,
whereas more complex shapes score lower.

The Groundino model filters for specific content within the
masks, such as recognizing specific materials or building parts.
Identifying windows and doors helped scale all images relative
to each other, assuming similar heights of windows and doors
in relation to their proportion and object count. This approach
allowed for comparability of diverse building types in the da-
taset, depicted in different proportions and perspectives, in a
quantifiable manner. Moreover, material usage could be linked
to compositional features, enhancing the analysis.

**RESULTS AND DISCUSSION**

Upon examination, Stable Diffusion’s generated images exhibit
notable visual complexity and precision, though not without
constraints. Approximately 15 percent of the depicted cities
manifest as generic representations, marked by abstract, mod-
ernistic forms, potentially attributable to data paucity. Another
15 percent of the renderings portray specific urban segments
without encompassing the full cityscape; for example, Vienna
is primarily represented through its 19th-century architectural
elements. Such selective portrayals raise concerns, especially
in regions associated with conflict, devastation, or vulnerable
living conditions. Overall, the generated imagery reveals persis-
tent data gaps and biases that preferentially represent dominant
architectural aesthetics, as noted more generally by Kate
Crawford’s analysis of AI perpetuating inequality. Architect
Andrew Kudless has highlighted similar inherent biases within
the Laion dataset used to train Stable Diffusion.

To maintain research integrity, we removed cities visually identi-
fied as misrepresentative. Hallucinated places can be visually identi-
cified using a tailored CNN similarity model. Specifically, we employed
the EfficientNetB0 model from TensorFlow Keras, trained on
ImageNet and optimized for image classification. When trained
on a new dataset, EfficientNetB0 detects specific patterns, ob-
jects, and features. Using this model, we identified and removed
cities similar to the generic modernist pattern.

With the specific case study of measuring carbon embodiments
through image segmentation, we wanted to demonstrate how
entanglements between form and synthetic data can offer
unique insights for real-world applications. Specifically, we ex-
amined how a compositional reading enables inferences about
the carbon footprints of places and cities. For this, we leveraged
the image segmentations in multiple ways. First, detected build-
ing parts like windows and doors were used to scale images
proportionally, enabling area comparison. Second, we tallied material usage per city based on derived from the visual material classification and construction assumptions. For instance, painted facade areas were categorized as either concrete or stone structures by assessing the scale to eliminate timber construction, and the dominance of horizontal or vertical windows. Third, we multiplied the accumulated material usage by typical embodied and lifecycle carbon measures for each, incorporating approximate error ranges. This allowed sorting, ranking, and visualizing places by their architecture-embedded carbon footprint. While approximate, this demonstrates using typological image analysis to computationally infer relative sustainability performance based on formal composition.

In detail, the initial image segmentation process operates at an architectural resolution, decomposing buildings into walls, roofs, openings, balconies, setbacks, and contextual elements including plantings. However, it only discerns material changes
like façade details, cornices, or smaller artifacts when contrast is highly pronounced. Practically, not all architectural features can be extracted from a single segmentation; material recognition and object detection require different search resolutions. A major advantage of the generated imagery is the exclusion of mobile elements like people, animals, furniture, or vehicles. Thus, any non-architectural content pertains solely to landscapes, skies, or vegetation.

The utilization of the Groundino model for segment classification presented significant challenges. Initially calibrated with imagery predominantly sourced from social media platforms, the model demonstrated notable difficulties in recognizing more abstract object categories not typically featured in everyday imagery. For instance, it exhibited an inability to identify common constructs such as ‘house’ or ‘building,’ despite their presence in the original segmentation. Similarly, painted walls were not categorized as ‘walls,’ but were identifiable through explicit color markers, such as ‘yellow.’ Consequently, we initially labeled detected colors as painted surfaces. Subsequently, an overlay with terms associated with building structures, such as ‘bricks,’ facilitated a more nuanced association of painted surfaces with their respective underlying material categories. This approach highlights the inherent complexities in object detection within image recognition, which employs a taxonomic scheme of categories. Such a scheme can lead to ambiguous classifications in compositional analysis. For instance, windows were frequently misclassified as paint material, owing to the model’s literal interpretation of reflections and transparencies in glass surfaces as the material they appear to be. Moreover, the model encounters difficulties in accurately depicting objects subject to significant perspective distortions, such as elements of a wall viewed along a street. This limitation appears consistently across various object classes and images. However, this issue did not impact our analysis focused on examining proportional relationships in feature distribution across different building types.

Considering all identified limitations, and the current stage of object segmentation model development in computer science, the methodology allows to identify quantifiable trends. However, these trends are not sufficiently precise to provide accountable metrics. It is crucial to note that such preliminary quantitative analyses render only general approximations and necessitate subsequent in-depth investigations. Preliminary findings suggest that the carbon footprints associated with specific building typologies may not align with the broader carbon emission metrics as proposed by other studies. Cities with diverse carbon footprints manifest globally, pointing to a myriad of underlying factors. Furthermore, visual representations such as heatmaps that display material usage patterns indicate that the correlation between material constituents and a city’s carbon footprint is not strictly linear. It’s noteworthy that cities with similar carbon footprints can exhibit stark contrasts in material compositions. In summation, this data analysis approach offers a granular perspective of environmental attributes, often aligning more closely with specific neighborhoods and projects rather than encompassing entire urban contexts, thereby offering a nuanced understanding.

CONCLUSION
Our research indicates a potential to distinguish a synthetic database based on their compositional attributes. Preliminarily, LLI models, specifically, in this study, the Stable Diffusion 2.1, demonstrate the capability to generate architectural scenarios with a diversity and realism mirroring the variability found in the built environment. This research principally serves as a proof-of-concept for the proposed workflow. Acknowledging the possibility of data misalignment, we posit with cautious optimism that our methodology facilitates a more nuanced analysis than conventional methods offer. Especially in pedagogical contexts, such a strategy can be instrumental in illuminating the multifaceted nature of urban landscapes and living patterns. The dataset generated, in itself, offers detailed insights into regions that might not ordinarily receive attention.

In closing, this research demonstrates the potential of using generative AI and synthetic data to computationally investigate urban architectural typologies. The complex entanglements between textual prompts, image generation, segmentation, and feature analysis showcase an emerging methodology for design studies. While requiring further validation and refinement, our workflow points toward new possibilities for mapping relationships between built form, and sustainability factors like carbon footprint through an AI-enabled computational lens. As the concept of ‘type’ remains a dynamic term within architectural discourse, it is essential to recognize that data-centric typologies should resist being anchored in static compositional weightings. Each machine learning model employed in this study presented its distinct advantages. We hypothesize that, contingent on scale and purpose, architects will craft distinct compositional feature models, underscoring the imperative for AI literacy in architectural domains. As generative models continue to advance in sophisticated and scale, they may unlock new urban analytical capabilities and inform data-driven yet humanistic perspectives on architecture and the city. A study centered on data, but approached from an architectural perspective, can offer a new granular perspective on environmental issues and their physical ties offering new interdisciplinary roads for architects. This research provides an initial foray into this promising cross-disciplinary space between urbanism, sustainability, and artificial intelligence.
Figure 5. Left: The top nine cities with a low carbon footprint: Sintra, Tiraspol, Monza, Sirjan, Toride, Najafgarh, Tecamac, Neliore, Setubai. Right: the top nine cities with a high carbon footprint related to architectural type: Almere, Hinthada, Hong’an, Hong Kong, Brahmaspur, Islamabad, Giradot, Hosur, Brisbane. Image by author.
ENDNOTES


