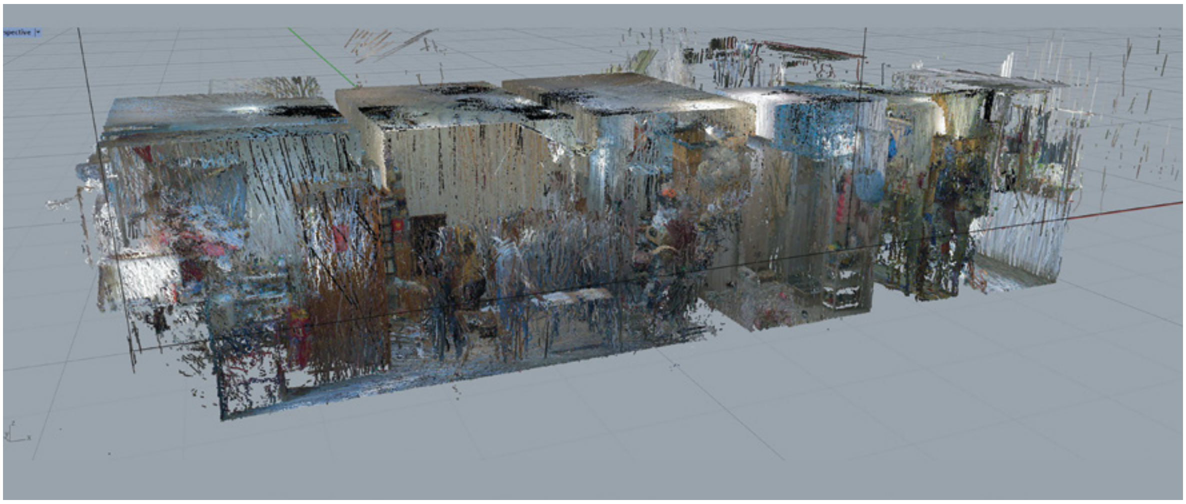
Generative-inferential spatial-semantic AI:

normative rational social imperatives at a new socio-technical horizon

GISS人工智能：新社會技術視閾下的理性社會規範

Daniel ELKIN, Gerhard BRUYNS 卓維德



01 A residential interior recorded using the Leica BLK2GO 3D Environment Scanner.

Environmental scanning technologies, such as light detection and ranging (LiDAR), photogrammetry, and simultaneous localization and mapping (SLAM), combine to describe existing architecture using point cloud and mesh geometry. Semantic segmentation, which classifies point cloud geometry into categories of real-world objects, is increasingly automated through large-scale computing models or artificial intelligence (AI). Similar to image-generation models like Midjourney, semantic segmentation could potentially be reverse-engineered to generate coherent point cloud geometry and new digital architecture from existing data. This paper reviews emerging technological advancements to support a position on their relationship to architecture and social change. It argues that, considering critical sociological theory, the technical horizon of generative-inferential spatial-semantic (GISS) AI should prompt architects to develop new technical competencies while re-centralizing normative social imperatives in practice and pedagogy. Drawing on critical sociological theorists Theodor Adorno and Max Horkheimer, and globalization theorists Leslie Sklair and David Held, this paper contrasts normative and technically rational responses to technological change.

環境掃描技術融合了「光探測與測距」(即激光雷達LiDAR)、「攝影測量法」(photogrammetry)、「即時定位與地圖構建」(SLAM)等感測方式，通過點雲及網格處理描繪現有建築架構。語義分割(Semantic segmentation)將真實環境物件以點雲網格進行分類，並通過大型電腦運算模型或人工智能提高自動化程度。類似於Midjourney等圖像生成工具，語義分割可以通過逆向工程生成連貫的點雲網格及全新數碼架構。本文將檢視新興技術發展與建築及社會變遷的關係，論證批判社會學理論及嶄新的GISS人工智能技術如何促使建築師在發展新技術能力的同時，將規範性的社會要求重新集中在實踐和教學法中。基於批判社會學理論家Theodor Adorno及Max Horkheimer和全球化理論家Leslie Sklair及David Held的思想，本文將對技術變革的規範性和理性反應進行對比。

Introduction

This article reviews developing digital surveying technology, and projects its development into a spatial generative AI. Given the implications this would have for the architecture professions, the article states a position on practice and pedagogy from the standpoint of critical sociological theory. Specifically, the paper will explain environment scanning tools' capabilities and discuss semantic segmentation workflows in site documentation. Following, the paper briefly overviews generative text-to-image AI tools to illustrate analogues between their generative structure and emerging semantic segmentation software. The article contends that generative AI tools designed to inferentially combine 3D scan geometry into new digital architecture are forthcoming: generative-inferential spatial-semantic (GISS) models which create new architecture from existing data will likely develop. The last section states a position on potential responses to this technological change from the standpoint of critical sociological theory and globalisation theory, especially dialectical comparisons between 'normative' and 'technical' rationalities. The article takes the position that, given that generative AI models for creating digital architecture will develop, there is a need to examine practice and pedagogy in the context of its social position in capitalist globalisation and to centralise social imperatives in architectural thought.

Are there parallels between semantic digital surveying and image-processing technologies that anticipate generative architectural AI?

Digital Surveying Technology: Photogrammetry and LiDAR

Digital surveying refers to technologies that create digital models of real-world physical environments for visualisation and manipulation.[1] They are increasingly common in construction and architecture, though they present barriers to entry, including equipment cost and training. Some digital surveying technologies are referred to as '3D scanning', and they offer promising benefits for architectural practice. Digital surveying can be compared to manual surveying, which uses tools familiar to architects and surveying professionals within different ranges of specialisation.[2]

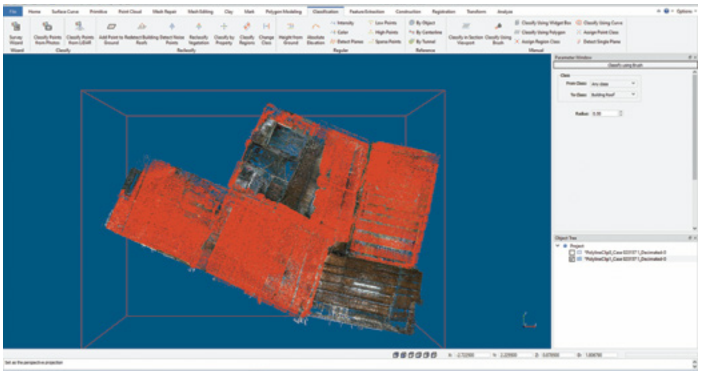
In comparison to other site survey techniques, 3D scanning is comparatively accurate, non-destructive, faster and safer, and creates digital environment models that are duplicable and transferable. There are two broad categories of 3D scanning technology, photogrammetry and LiDAR-based, which are often combined. Photogrammetry interpolates a visually and/or dimensionally accurate digital model from numerous photographs,[3] while LiDAR scanning projects infrared laser light to locate physical objects, compiling measurements into a digital model.[4] Many 3D scanning tools combine photogrammetry and LiDAR sensing to create full-colour, dimensionally accurate digital site documentation.

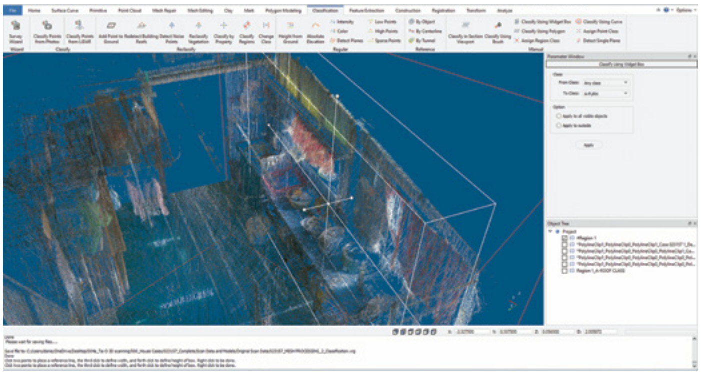
Digital surveying technologies offer powerful possibilities. For example, the Leica BLK2GO portable platform records 420,000 points per second with a dimensional precision of +/- ten millimetres after a two-minute scan duration, and operators using the equipment can increase scan accuracy by conducting longer scans.[5] Depending on the platform, 3D environment scanning can create a highly accurate, full-colour digital model of an environment by simply walking through it for a short time [01]. Most technologies use proprietary software to process the digital geometry they gather. Photogrammetry-based scanners mostly create digital mesh models, while LiDAR-based scanners create point clouds, which require greater computer processing power to view.

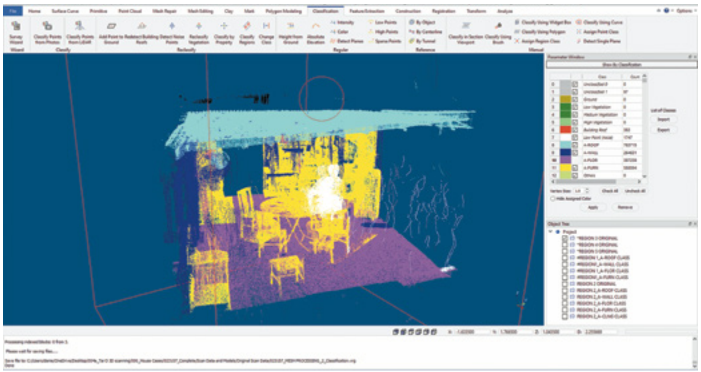
Despite these challenges and the technologies' often significant investment cost, they offer obvious practical benefits to architectural practice. Accurate site surveys are critical to design and drawing production, and as 3D scanning technologies become cheaper, they will help architects lower the time costs associated with accurate site surveys. The technology also supports digital conservation and digital humanities research.[6] However, technologies that create point cloud geometry often require an additional workflow, called semantic segmentation, to be useful, as discussed below.

Semantic Segmentation: Assigning meaning to coordinate regions

LiDAR-based digital surveying tools create 'point cloud' geometry. A point cloud stores the digital version of an environment using Cartesian coordinates (XYZ) and colour data (RGB) values.[7] Visualizing and manipulating space using point cloud geometry can be challenging. Leica's BLK2GO and Leica's other laser scanners require the proprietary Leica REGISTER 360 program to import scan geometry and convert it to other file formats.[8] Even short scans can create large files, with a three-minute scan creating a 500 Mb file and file sizes increasing by approximately 100 Mb for each additional minute.[9] Scanning stilt house residences in Tai O Village, for example, created some point cloud files larger than 2Gb, which can only be reduced using specialised workflows.[10] In comparison to mesh


02


03


04

02 Semantic segmentation using a 'brush'...
03 ...and a 'widget box'...
04 ...to segment a point cloud into classes.

image source: the Authors

or Non-Uniform Rational B-Splines (NURB) models, point clouds do not distinguish their content volumetrically or as finite elements: other formats store geometry data as surfaces and/or volumes which users can visualise and manipulate analogous to discrete objects, which cannot be done directly with point clouds.

Assigning point cloud content into regions analogous to real-world objects is called 'semantic segmentation.' A user or algorithm assigns points to 'classes,' following standards developed by the American Society for Photogrammetry and Remote Sensing (ASPRS) or other surveying standards.[11] Semantic segmentation is similar to assigning layer content in AutoCAD: a user selects content in a point cloud file and assigns it to a class – segmentation - based on its real-world meaning – semantic significance. Semantic segmentation software supports the interpretation of point cloud geometry, so a user needn't always select points manually or individually.[12] As a simple example, VRMesh, a program designed to create mesh geometry from point clouds, includes tools that compare vector angles and distances between points or relative point colour values to more rapidly designate point cloud regions as flat surfaces or features [02,03]. Point cloud semantic segmentation is important for communication and geometry refinement. Point cloud files inevitably include 'noise,' which can be eliminated if a region is known to include undesirable points [04].

How might the profession, if construed as a socio-technical field, respond to this disruptive possibility?

Generative AI: Inferential language, image, and spatial models

Given that this is a labour-intensive workflow, many researchers have developed AI tools to expedite semantic segmentation.[13] In this case, artificial intelligence refers to large-scale, neural network statistical inference models. Statistical inference models depend on large bodies of data to infer or 'guess' the semantic significance of data.[14] The premise for applying inferential AI to semantic segmentation is similar to, though more sophisticated than, tools in VRMesh: given large bodies of user-classified semantic segmentation data, the software can infer which parts of a point cloud correspond to which real-world objects. With human verification, AI models for inferring semantic segmentation accelerate the workflow. If examined in the context of generative AI tools for image generation, however, it seems feasible that this analytical workflow could be reversed into a spatial generation tool.

ChatGPT, Midjourney, Stable Diffusion, DALL-E Open AI, and other generative AI platforms or tools depend on large-scale, neural-network inference like AI semantic segmentation tools, but in reverse. Prior to text-to-image generation AI platforms, AI development focused on semantic image description programs designed, for example, to give a text caption that describes an image. Neural networks are machine learning programs that identify data and make decisions using a model-like animal brain structure. A neural network programmed to assign text captions to images, must first represent language and common-sense knowledge in addition to object recognition, to identify important semantic pieces within an image, and produce a coherent description of an image's overall content,'.[15] Increasing a neural network's accuracy requires 'training' the network based on a large body of data, asking it to identify image content and correct mistakes. An important development in computing required for *generative* AI was the development of generative adversarial networks (GANs). A GAN is a structure in which 'Image generation' is driven by two antagonist neural networks. New images are created by the first, known as the generator, and are contrasted with a training set of samples by the second, known as the discriminator. The generator and discriminator iterate images through 'diffusion,' which entails successive addition of pixel noise to an image, testing of the image by the discriminator, and revision of the image by the generator. [16]

Differences between a general explanation of generative AI and the programming needed to make one work are not trivial. However, given semantic segmentation is seeing automation using a similar GAN structure,[17] it is possible to anticipate a diffusion-based generative model that, rather than iteratively making *images* using *pixel noise* for a GAN to test, could iteratively form *digital spaces* using *point/voxel noise* in the same way. This would create a generative-inferential spatial-semantic (GISS) AI model. Along with the extensive work required to program a GISS, researchers would require an extensive body of 3D scan data to train it, which may be a long time coming. Still, it is possible to anticipate GISS platforms' emergence, enabling a broader-based generation of highly detailed digital spatial models from prompts. The architectural professions might then see the same disruption seen in other creative fields when AI platforms became broadly available.[18] Considering this possibility, the last section of this paper states a position on possible responses from people practising and teaching architecture.

A Position on technical rationality: innovate and compete, centralise social imperatives

In this section, the paper takes a position on the technological development described above based on 'critical theory,' sociological texts, globalisation theory, and the implications sociological research has for architecture's response to new technology as a profession. Critical theory 'emphasises the critical, reflective use of reason in assessing and advancing society's implementation of Enlightenment values' and is inherently normative, i.e., prescriptive and non-neutral toward societal change. Critical theory texts emerged from the Frankfurt School during the mid-to-late 20th Century, including *Dialectic of Enlightenment* by Max Horkheimer and Theodor Adorno.[19] While inherently Eurocentric, the premise of 'normative' and 'technical' rationalities bear discussion during seismic socio-technical changes.

In summary, Horkheimer and Adorno point to the emancipatory promises of the Enlightenment: humanistic, personal liberation and satisfaction of individual and collective purpose, which have failed to materialise when, in part, societies respond to technology in ways that are technically rational rather than normatively rational. Technical rationality 'focuses on the efficiency of the method used to accomplish goals, without any thought given to the ethics of the goals being pursued and the values they serve.' Conversely, normative rationality entails '[the use of] reason in pursuit of goals that advance enlightenment norms or values of social equality, democratic participation, and human flourishing, as opposed to...instrumental, strategic, or technical control and domination' .[20] Put simply, technically rational responses extort what societies using technology *can* do, while normatively rational responses emphasise what they *should* do.

Without question, this literature can support reactionary positions toward new technology. It bears stating that disruptive technology has emerged accompanied by calls to 'innovate or perish,' [21] for a reason, as far back as David Celento's writing in 2010.[22] While architects work to innovate and compete, it also bears the statement that architecture is a *socio*-technical profession that should remain cognizant of relationships between new technology and contemporary social imperatives. Globalisation theorists Leslie Sklair and David Held theorise that climate change and class polarisation are crises *endemic* to capitalist globalisation: problems which the socio-economic world system we live and practice in cannot solve because it *depends* on inequality and ever-increasing consumption.[23] Exploring generative AI's promise, architects must also educate themselves on the technology's implications for body politics,[24] education,[25] and the promotion of implicit biases.[26] As an open question subject to research and worthy of architects' attention, the relationship between generative AI development and symptoms of the global capitalist crisis, including the commodification of creativity, extension of consumption culture, and concentration of wealth via technology, should be examined.

Conclusion

The position taken here describes how recent developments anticipate generative-inferential spatial-semantic (GISS) AI, a technology which could generate new digital architecture from large bodies of 3D scanning data. These current and emerging technologies have evident value for architecture as an evolving profession and, despite their potentially disruptive effects, should be integrated into architects' technical knowledge. Concurrently, the paper argues that a socio-technical standpoint in architecture, informed by critical and globalisation theories from sociology, projects a resounding need for architects to remain cognisant of global social imperatives in their work, education, and critical thought.

Notes

1 Rashed, T., & Juergens, C. (2010). *Remote sensing of urban and suburban areas*. Springer.

2 Nadolinets, L., Levin, E., & Akhmedov, D. (2017). *Surveying instruments and technology*. Taylor & Francis, CRC Press.

3 Kasser, M., & Egels, Y. (2002). *Digital photogrammetry*. Taylor & Francis.

4 Wang, C.-C., & Thorpe, C. (2004). Simultaneous localization, mapping and moving object tracking. Thesis (Ph.D.)--Carnegie Mellon University.

5 Leica Geosystems. (2024). *Leica BLK2GO - Handheld Laser Scanner*. Leica BLK2GO.

6 Valente, R., Oreni, D., Barazzetti, L., Roncoroni, F., & Previtali, M. (2019). Documentation of historical architectures through the combined use of digital surveying techniques. *2ND International Conference of Geomatics And Restoration (GEORES 2019)*, 42(2), 1121–1125.

Chen, S., Yang, H., Wang, S., & Hu, Q. (2018). Surveying and digital restoration of towering architectural heritage in harsh environments: A case study of the Millennium ancient watchtower in Tibet. Sustainability, 10(9), 3138-.

Hatzopoulos, J. N., Stefanakis, D., Georgopoulos, A., Tapinaki, S., Pantelis, V., & Liritzis, I. (2017). Use of various surveying technologies to 3D digital mapping and modelling of cultural heritage structures for maintenance and restoration purposes: The Tholos in Delphi, Greece. *Mediterranean Archaeology & Archaeometry*, 17(3), 311–336.

7 Zhang, G., & Chen, Y. (2022). *Towards optimal point cloud processing for 3D reconstruction*. Springer.

8 Leica Geosystems. (2024).

9 L. Lai, personal communication, September 6, 2024

10 Elkin, D., Leung, C.-Y., Wang, N. X., & Yan, W. Y. (2023). Digital cultural heritage conservation: sampling stilt houses in Tai O Village. In Proceedings of IASDR 2023 [Short Paper]. *The 2023 IASDR Congress*, Milan, Italy.

11 American Society for Photogrammetry and Remote Sensing (ASPRS). (2014, March). ASPRS Accuracy Standards for Digital Geospatial Data.

12 Bearman, A., Russakovsky, O., Ferrari, V., & Fei-Fei, L. (2016). What's the Point: Semantic Segmentation with Point Supervision. In *Computer Vision - ECCV 2016, PT VII* (Vol. 9911, pp. 549–565). Springer International Publishing.

13 Sehar, U., & Naseem, M. L. (2022). How deep learning is empowering semantic segmentation: Traditional and deep learning techniques for semantic segmentation: A comparison. *Multimedia Tools and Applications*, 81(21), 30519–30544.

Jiang, S., Li, G., Yao, W., Hong, Z., & Kuc, T. (2020). Dual pyramids encoder-decoder network for semantic segmentation in ground and aerial view images. Copernicus GmbH.

Li, Y., Shi, T., Zhang, Y., Chen, W., Wang, Z., & Li, H. (2021). Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. ISPRS *Journal of Photogrammetry and Remote Sensing*, 175, 20–33.

14 Gillain, E. (2024). *Demystifying Artificial Intelligence : Symbolic, Data-Driven, Statistical and Ethical AI*. (1st ed.). Walter de Gruyter GmbH.

15 Hanafy, N. O. (2023). Artificial intelligence's effects on design process creativity: "A study on used A.I. Text-to-Image in architecture." *Journal of Building Engineering*, 80, p.5.

16 Hanafy, N. O. (2023).

17 Jiang, S., Li, G., Yao, W., Hong, Z., & Kuc, T. (2020). Dual pyramids encoder-decoder network for semantic segmentation in ground and aerial view images. Copernicus GmbH.

18 Minevich, M. (2024, May 9). The Dawn of AI Disruption: How 2024 marks a new era in innovation. *Forbes*.

19 Horkheimer, M., & Adorno, T. W. (2020). *Dialectic of enlightenment*. Stanford University Press

20 Dillon, M. M. (2020). *Introduction to Sociological Theory: Theorists, concepts, and their Applicability to the Twenty-First Century*. p.183-184

21 Kahn, E. (Edward A.). (2007). *Innovate or perish : managing the enduring technology company in the global market*. Wiley.

Lugo-Morin, D. R. (2022). Innovate or Perish: Food Policy Design in an Indigenous Context in a Post-Pandemic and Climate Adaptation Era. *Journal of Open Innovation*, 8(1), 34-.

22 Celento, D. in Corser, R. (2010). *Fabricating architecture: selected readings in digital design and manufacturing* (1st ed.). Princeton Architectural Press.

23 Sklair, L., & Sklair, L. (2002). *Globalization : capitalism and its alternatives* (3rd ed.). Oxford University Press.

Held, D. (2004). *Global covenant: the social democratic alternative to the Washington Consensus*. Polity.

24 Wodziński, M., Rządeczka, M., Szuła, A., Sokół, M., & Moskalewicz, M. (2024). Visual Stereotypes of Autism Spectrum in DALL-E, Stable Diffusion, SDXL, and Midjourney. *arXiv.Org*.

25 Hwang, G. J., & Chen, N. S. (2023). Editorial Position Paper: Exploring the Potential of Generative Artificial Intelligence in Education: Applications, Challenges, and Future Research Directions. *Educational Technology & Society*, 26(2).

26 Allen, L., Xu, W., Nishikitani, M., Patil, V. A., & Bradley, D. (2023). AGE BIAS IN ARTIFICIAL INTELLIGENCE (AI): A VISUAL PROPERTIES ANALYSIS OF AI IMAGES OF OLDER VERSUS YOUNGER PEOPLE. *Innovation in Aging, 7* (Supplement_1), 986–986.

Daniel ELKIN holds two Masters of Architecture and is an Associate Professor at The Hong Kong Polytechnic University School of Design. He researches housing, digital surveying, and social morphology.

Daniel ELKIN 擁有兩個建築學碩士學位，現職香港理工大學設計學院副教授，專注於住房、數碼測量及社會形態學的研究。

Dr.ir Gerhard BRUYNS is an architect, urbanist and Associate Professor at The Hong Kong Polytechnic University School of Design, where he researches morphology and super-tall, super-dense settings.

卓維德博士 為建築師暨城市規劃師 ，亦是香港理工大學設計學院副教授。卓博士的學術研究集中於形態學、超高層及高度密集的建築設計。