The International Journal of the Image ISSN: 2154-8560 (Print), ISSN: 2154-8579 (Online) Volume 16, Issue 2 https://doi.org/10.18848/2154-8560/CGP/v16i02/55-77



Original Research

Architecture and the Built Environment Seen Through the Eye of the Lens and AI: Does It Matter Which?

Martine Hamilton Knight, Nottingham Trent University, UK

Received: 10/10/2024; Accepted: 03/26/2025; Published: 06/18/2025

Abstract: Every day, we encounter photographs that have been manipulated via computer software. However, there is an argument that deceit has always been the case with photography, and audiences of such pictures are expected to be passive in acceptance of their controlled and constructed content, regardless of provenance. However, until recently, human agency has been central to physical creative processes. Using architecture as subject focus, this article questions whether creative AI software programs can mimic known built environments effectively and if Al-generative images embodying "photographic" accuracy may be aligned with the actual places and spaces they depict. Audiences of architectural photographs should be able to "read" buildings and are generally less concerned with the method of image creation. This suggest that Al-generated works pose no greater threat to the perception of "reality" than signifiers of the subject produced by conventional cameras. Using the software program Midjourney, this study employs Methodological Pragmatism, including image analysis, interviews, and gualitative research to question if creative photorealistic AI text-to-image outputs may provide a trusted view of the world. Viewer cognition and interpretation are discussed, along with the type of authority exposing architecture as subject, asking if audience "faith" has been taken away since the emergence of AI images? The research concludes that if we could understand what we are looking at, signifier or signified, it rarely matters whether a picture was created by a camera or by a series of text-to-image prompts. Education and trust are vital. As always in photography, context is king.

Keywords: Al-Generative Image, Architecture, Neuroscience, Photography, Text-to-Image

Introduction

Traditionally, photography has often been considered the medium par excellence that is able to register one particular space at one particular moment, once and for all. (Van Gelder 2009, 77)

The premise of this work is that photographic images (within the broad caveats of subjectivity) could convey physical spaces in a manner that can align to our understanding of being there. While a picture is often a poor substitution for one's actual presence in a space or place, it is the medium of photography that is most employed as an indexical referent to the subject depicted. As Victor Burgin wrote in 1987, "More than any other textual system, the photograph presents itself as 'an offer you can't refuse' " (Burgin 1987, 146). He expands



on this, describing the photograph and its subject matter as a "what is it?" puzzle piece, where, once understood, becomes transformed, invested in to the degree of acceptance and adoption. "A fact of primary social importance is that the photograph is a *place of work*, a structured and structuring space within which the reader deploys and is deployed by what codes he or she is familiar with in order to *make sense*" (Burgin 1987, 153).

Nevertheless, much as digital photography has replaced analog as the ubiquitous means of picturing the world, it has come to pass that this most popular method of capture globally is now fighting for airtime and page space with generative AI images. Alongside this change, our capacity for discerning provenance is rapidly being contested. In 2025, we are subject to a variety of image production means. These include solely lens-based capture using traditional analog and digital cameras, together with "blended images" created by physical cameras but with the addition of postproduction retouching using popular computer programs such as Adobe Photoshop or Serif Affinity. This retouching can be facilitated manually by an editor using "clone tools," which pick up pixel data from one area of a photograph and place it elsewhere. In addition, software packages such as Adobe Photoshop now include the expanded ability for a user to employ their Generative Fill action, which draws on its own image dataset to create replacement content for parts of a photograph that an editor wishes to replace (Adobe 2025). The third method of picture-making is using creative AI text-to-image generation software programs such as Stable Diffusion, Dall-E, Adobe Firefly, and Midjourney.

This article proposes that provided we *understand* the context of what we see, Algenerative images pose no greater threat to our acceptance of "reality" than signifiers of the subject produced by analog and digital cameras. Using the architectural realm as a subject focus, the work questions why we challenge machine-generated visions as a trusted view of the constructed world. Architecture is a useful territory to make these investigations as: Whichever means of architectural image-making is employed, viewers of these images and those who are familiar with the medium of architecture will concur that "semiotic understanding of language, communication and identity has brought us virtually to the point where the language of architecture and those of advertising and the media cross over into one another's territories" (Coates 2012, 11).

A key question is whether creative AI-generative image programs are currently able to mimic the known built environment effectively for use in professional documentation that reports on conditions at a given site, or whether lens-based capture still prevails. For viewers, is the fear that they have had viewing "confidence" taken away from them, "hoodwinked" by photorealistic AI? Nicholas Mirzoeff (2015, 73) states, "Seeing the world is not about how we see but about what we make of how we see." With AI imagery, we are still grappling with the rules of navigation, affecting our cognitive and emotive resonance when it comes to reading and accepting such works.

Methodology

In addition to literature review and auto-ethnography of lens-based and computational photography, ethical approval was obtained to carry out interviews with specialists in AI image production and the field of architectural neuroscience. The respective contributions of US architect and academic, Damon Leverett, and Brazilian (but Italy based) architect and neuroscience specialist, Clarissa Machado have exceeded the scope of conventional research participation. Leverett's use of Midjourney (2024) as the chosen creative AI text-to-image generative software for this work brings insight to the subject of architectural capture. Machado's neuroscientific understanding of the emotive response of viewers to the built environment and its imagery is empirical to these findings in understanding audience confusion over image provenance and content.

The Foundation of Images

Few would contest that photography has the power to manipulate and be manipulated. It has always been the case with analog capture, long before digital cameras and their associated postproduction software came into play. Image makers hold power over their construction in camera—lens choice, viewpoint, depth of field, and so on, to place emphasis in a story or to change the spatial relationship between objects in a scene. The same applies in postproduction, there is a history of manipulation from Oscar Rejlander and Henry Peach Robinson onward to shape their narratives—a compelling argument that deceit has always been the case with photography (Hacking 2012, 113) (Figure 1).



Figure 1: A Composite of Over 30 Negatives—The Two Ways of Life—Oscar Gustave Rejlander (1857) Source: Princeton University Art Museum

Why, in general though, do we accept photographs, and in particular, those depicting the built environment, a genre rarely contested in the same manner as advertising images might be—as trustworthy? In The Life and Death of Buildings, Joel Smith (2011, 61) warns of the ability of architectural photographs to "change" reality, "By its presence alone, the camera modifies the shape and proportions of the space around it."

Examination of socio-technical constructs of photography have been explored in the context of fine art practice and cultural object, together with its status as commodity (Cruz and Meyer 2012). Meyer (2005, 10) had earlier described professional photographers—and in the case he cited—photojournalists—as, "part of a communication regime. The members of this regime are part of a shared social network, as are most people in workplaces, but in addition, the nature of their work is highly coupled to the communication of visual information." In the same manner, this author is a professional architectural photographer of long-standing, resisting the use of ultra-wide-angle lenses with their distorting outcomes, the temptation to employ theatrical lighting, or deviances in color temperature to create effect. The medium becomes a delivery vehicle for her audience for architectural photography is generally a learned one; architects or building professionals in the main, with a desire to understand the age and design of the subject itself. The challenge is to convey a space in the way it might look and feel to a visitor. The work is about the architecture, less the photography.

Even for viewers who are not "in the industry," the aim of architectural photography should be to give a sense of why a building exists in the form it does and how users navigate its spaces for work, rest, or play. Conveying depth and three-dimensionality through viewpoint, lighting, and use of models is very much the tried and tested workflow for many practitioners of the genre.

Several photographs (taken by the author of this study) of historic architecture and monuments in *The Buildings of England* series of guidebooks founded by Sir Nikolaus Pevsner in 1951 and held in academic libraries globally, underwent extensive computer manipulation in the postproduction stage. (Yale University Press 2024). Removing gum and litter on paving stones and sidewalks, together with people, retailer branding, and badly parked vehicles were but a few of the unwanted visual elements redacted, but these actions were achieved *prior* to the release of 2022's Photoshop AI-Generative Fill. The results facilitated a reductive tableau to communicate this subject to viewers with clarity.

Partly this is established by the need to examine each building in isolation as a period case study of its architectural form, and partly because there needs to also be an authoritative visual document of each structure that is not judged by the date or time in which it was recorded. (Hamilton Knight 2022, 86) (Figures 2–5).

HAMILTON KNIGHT: ARCHITECTURE SEEN THROUGH THE EYE OF THE LENS AND AI



Figures 2 and 3: Before and After—Cars and People Edited Out. Lichfield Cathedral (Wakeling and Pevsner 2024) Source: Hamilton Knight 2022



Figures 4 and 5: Before and After—Signage and Shop Fittings Edited Out. Governor's House, Newark (Hartwell et al. 2020) Source: Hamilton Knight 2022

This methodology is decisive human agency, which involves computers but unlike computational photography, where automated algorithms create crowd-pleasing outcomes, such as we see with our smartphones, it is instead, long form, laborious, and exacting. However, now there is the opportunity to circumvent this creative labor, and bring in machine learning, facilitating responses via text-based prompts. These are formed via generative software programs founded on extensive image datasets.

AI images are the source of much debate, and the topic of conferences, publications, inquiries, and policymaking. There was a "honeymoon" period of 2021–2022, when visual outcomes were largely in the realm of improbable, exhibiting a style that was discussed at the Royal Photographic Society's conference on Photography and Artificial Intelligence (Hamilton Knight 2023). Such pictures allowed "the viewer satisfaction that these works are fantasy, a 'never have been, or will be,' as opposed to Roland Barthes' (1977, 44) descriptions of traditional analog photographs with their 'here-now' or 'here-then'-ness."

At the time, there were gaps in the software's ability to mimic certain forms, but these were publicly contested and rapidly refined to greater and greater sophistication, to a point where we now struggle to define what is lens-based origin and what is not. It is the real crux of this inquiry and where audience mistrust and confusion may prevail. Can AI recreate an actual location, and therein, with enough photographic accuracy to replace human endeavor with lens-based media? Moreover, is an AI-generative image any less "real" than a photographer's manipulated lens-based outcomes?

This work aims to find out whether a program can accurately create its own example of an architectural project that a human with a camera would conventionally otherwise document and manually refine in postproduction. It asks whether AI text-to-prompt images are capable of embodying "photographic" accuracy and alignment to the built environments they depict? If so, can a client avoid sending a photographer to capture something which AI could generate for them instead and would an audience realize? Let us now turn to these images and their form when applied to architecture.

Datasets and Prior Learning

Firstly, a baseline needs to be established regarding picture datasets from which generative AI software creates its outcomes. These are limited by their specific content and scale, and we should confirm that such programs are only able to make results based on prior learning. They do not possess the ability to invent image forms that have no basis outside the training data that has been imported. While their directories continually expand (a product of saved image results from users' text prompts and other image additions), datasets can still only occupy their new ground based on existing picture content in their systems.

This means that they only have limited scope when tasked to mimic specific architectural styles. If their directories do not contain sufficient information to identify and replicate nuances within a time period or design rationale—for example, Victorian Gothic—they

cannot be expected to respond with accuracy in their image returns. In his *Teaching the Machines*, Milo Keller (2021, 6) states, "Machine learning is neither neutral nor truly autonomous. Neural networks only know the part of the world they are given to see and are therefore by 'nature' conservative and biased." Furthermore, Neil Leach (2022, 91) writes in his book about architects working with AI, that "the gaze of the architect is not neutral. It has been trained, no less than a neural network has been trained."

So, from this, we can say that in as much as a dataset is restrictive, it only knows what it knows, our human image-making and viewing is also shaped by our frames of reference as well as our learning and objectivity. An architecturally uninformed photographer in front of a castle may only choose to photograph it in a "generic" way, unable to "see" the specifics of a style. Such caveats may pose similar limitations to those of a dataset, nonetheless, despite the photographer not understanding that they are "seeing," lens-based media will still pick up the detail, something a viewer with knowledge e.g., an architect, could then "read."

Making Architectural Images

Can an experienced user of a creative program such as Midjourney guide an architectural image return that is controlled and accurate? Damon Leverett cites the "Spanish Mediterranean" as a type that many large programs can model using their thousands of imported photographic training examples. Where the style prompt has a considerable library to draw upon as Midjourney does, user keyword inputs may be succinct and the results confident and assured (Figure 6).



Figure 6: Prompts for this Image Given to Midjourney V6, 2024: Photorealistic, Spanish Mediterranean Revival Style, Office Building. https://alpha.midjourney.com/jobs/18632c81-ee2d-4db4-9bc4-286b956f8271?index=0 Source: Damon Leverett

Outputs deemed "successful," meet with criteria known as "AI Alignment" (Lutkevitch 2023). Programs work on three levels to achieve alignment in their results. One is "Intended Goals"; those which meet the user's objectives, a hypothetical outcome that satisfies. Another is known as the "Specified Goals," which are programmed into the system's own objective functions and dataset. The third is the "Emergent Goals," which are those created as the system it serves grows and advances. We will not hit "save" when either Outer or Inner Misalignment occur. These errors may take place when mismatches between our own failure in identifying successful keyword prompts cause the software to return results that do not satisfy (outer misalignment). Conversely, inner misalignment may happen when there is a mismatch between what is written in the code and what the system advances.

Notwithstanding deliberate or unintended in adherence to legal, moral, or ethical codes designed into any given software by a user, there is also a gray area in the workings of AI to consider. Most systems function as black boxes, and it is impossible for human programmers to see inside the actual workings to observe and monitor the computational processing taking place within. While it is possible to change user text prompts driving results, the calculations taking place within a program such as Midjourney whose image dataset was trained on Laion's 5.8 billion image-text pairs, are subject to change each time they are performed (Beaumont 2022). Exact repeats in function are not fully controllable, the machine workings are its own. If however, an image dataset is compiled from a specifically curated image-bank, it is logical to assume a tighter parameter around outputs and black box "control" may become more equitable for users. Elcott and Trombley (2025) worked with approximately 2,400 images from two photographers to test this notion and found by restricting the dataset, they were able to "introduce *latent specificity* into the model with changing its underlying architecture or data."

Findings

Leverett conducted his own test with Midjourney V6 Alpha (2024) to evaluate how the software would fare when commanded to create photorealistic built environments. The division must be drawn here between using creative AI-generative programs such as Midjourney compared to software such as Adobe Photoshop and its "generative fill" command. The latter provides AI-generated content for use in *parts* of lens-based photographic images where the user wishes to remove sections of the picture—an unwanted tree, or parked car for example—and seeks alternative subject matter to replace it with. This approach today might be an appropriate methodology to be applied to Hamilton Knight's (2022) *Buildings of England* work, which at the time of creation in 2020, was the product of the author's manual Photoshop cloning. In many cases such as hers, these are the user's own images and intellectual property. Therefore, there is a large degree of control over the construction of the original image and the parts within it that are being reinterpreted by the software. However, this investigation is focused on using purely creative programs, which are making content afresh for use across the complete image file and specifically here, Leverett's tasking with Midjourney.

With the aid of Chat-GPT (OpenAI, n.d.) to create a storyline, Leverett compared the performance of a twenty-word prompt with a hundred-word and ultimately, a thousand-word prompt to see where the greatest control over results could be asserted.¹ To try and provide a parallel concept, if we think of many items found in our regular lives, we may associate increasing sophistication and product refinement based on a consistent foundation, e.g., a modest four-bedroomed home may share the identical floorplan, room layout, and physical geology of its plot with that of an "executive" counterpart. We would simply anticipate an elevation in scale and quality of materials, fittings, and with it undoubtedly, cost for the executive home. Could we also assume that enriching a series of text prompts in length and detail would provide responses where increasing sophistication and complexity ensued?

Not in Leverett's trial. All three outcomes broadly follow the same narrative as twilight scenes and plausibly within the same city district; specified as being in downtown Bangkok where a traditional tea-house stands adjacent to a modern high-rise tower. However, what is interesting is that each return does not utilize an identikit foundation image with layering of ever-increasing and more nuanced detail that would mimic an increasingly sophisticated directive in the text-prompting. Here, differences between the twenty-word (Figure 7) and hundred-word (Figure 8) outputs in intended goals are subtle. Leverett mostly discerned errors in the human figures depicted rather than the built environment depicted.



Figure 7: Output from Twenty-Word Prompt Given to Midjourney V6. 2024. https://cdn.midjourney.com/d8f54742-16e7-4220-8fc5-40b6bcd4ad48/0_0.png Source: Damon Leverett

¹To read the full series of prompts used by Leverett for this exercise, please see Leverett, Damon. 2024. "Interview by Martine Hamilton Knight." July 7. https://doi.org/10.5281/zenodo.14926151.



Figure 8: Output from Hundred-Word Prompt Given to Midjourney V6. 2024 https://alpha.midjourney.com/jobs/7c1abcc9-06ad-40b0-ad36-e3936cc8ebbf?index=0 Source: Damon Leverett

The thousand-word prompt (Figure 9) drives the most satisfying result for seemingly photographic quality rendering but still leaves some of the detail given in the prompts aside, the tuk-tuk seen in the first two images being one such element. Leverett (2024) has found that right across his work, word choices used in prompts will affect how the software responds, stating, "Clearly, what type of word is used is a factor, such as whether it is a noun, verb, or adjective."



Figure 9: Output from Thousand-Word Prompt Given to Midjourney V6, 2024. https://cdn.midjourney.com/44fda69a-65f3-4161-97a0-d24aab902dc8/0_0.png. Source: Damon Leverett

Leverett's findings fall into line with a 2023 study conducted by Joern Ploennigs and Markus Berger (2023, 6). They analyzed how architecturally focused users of Midjourney software apply their prompts via Chat-GPT, analyzing a dataset of 101 million queries to understand if, and how, those practitioners queried specific architectural concepts. They identified a known issue with large language models such as Chat-GPT, which make up missing information, something commonly referred to as "hallucination." This "shows that it is not that reliable on naming architects or examples for less well documented styles and will hallucinate and invent names including biographies of those architects." Nevertheless, they do "also confirm that Midjourney has a robust knowledge of many architectural styles, with a clear bias towards the kinds of styles that are of specific interest to its users" (Ploennigs and Berger 2023, 20).

They found that most users "do not come-up with perfect queries from scratch but normally develop them over multiple iterations by selecting the best variants or adding more terms (especially style terms)" (Ploennigs and Berger 2023, 18). In the current study, Leverett also confirms this matter, "As a heavy user of text-to-image tools, I am generally satisfied with how AI meets my goals. However, it takes many iterations to develop satisfactory results. The hard work is worth it." It is worth noting that the time taken for AI to create its return is fractional compared to the Herculean (albeit calculably and exactingly accurate) number of hours needed if he were to work with a CGI program for similar renderings in his architectural practice.

Consistency in Returns as a Basis for "Proof"

So far, we have confined the discussion to images exampling architectural styles and while it is clear that with the right circumstances, AI can be used to generate fairly authoritative textto-image examples based on popular themes, it is important to understand whether the software is able to mimic photographs of known buildings or locations, based on its understanding of what that place or space actually looks like.

To investigate this, Leverett was challenged to create an AI "photograph" of a building at a known address, set in our time. The issue appears to be that AI simply does not, or more likely, cannot approach requests of this nature by providing a response from actual examples in its datasets reflecting a specified place. Leverett used a random address, which he asked Chat-GPT for (OpenAI, n.d.). It came up with 100 Market Street, Sydney, NSW, 2000 Australia (Figures 10–12). From Google Streetview (2024),² it is clear to see that this address is midway along a city block of high density, high-rise modern buildings.

² https://www.google.co.uk/maps/place/100+Market+St,+Sydney+NSW+2000,+Australia/@-

 $^{33.8706103, 151.2062805, 1255} m/data = !3m2! 1e3! 4b1! 4m6! 3m5! 1s0x6b12ae3 fa81a93 bf: 0x53d3065 c8f8 fa359! 8m2! 3d-33.8706148! 4d151.2088554! 16s/g/11f3_trlt4! 5m1! 1e3? entry = ttu&g_ep = EgoyMDI0MDgyMy4wIKXMDSoASAFQA$



Figures 10, 11, and 12: Google Street View of 100, Market St, Sydney, NSW, 2000 Australia, Shown from Three Angles Along the Linear Frontage. The Building Appears to be a Mixed-Use Retail and Commercial Development. *Source: Google Streetview 2024*

However, Figure 13 reveals a very different street scene. According to Midjourney, 100 Market Street is sited on a junction. The actual street level is cropped out of shot and the architecture looks to be an eight-story stone faced, mixed-use block of the type favored by city designers at the close of the nineteenth century. Mature trees are in view and newer, high-rise towers lie beyond.



Figure 13: Prompt given to Midjourney—100 Market St, Sydney, NSW, 2000 Australia.—ar 16:9—seed 301—v6. 2024. https://cdn.midjourney.com/e7566b6d-f4d6-4359-9b19-7589496debaf/0_3.png Source: Damon Leverett

This is simply an approximation of what might be sited on such a city block, based on what Midjourney understands a typical city center could look like; a mixed-age mishmash of architectural responses, developed across several decades. It is not a report of the current built environment at 100 Market St, Sydney. As Celina Lei (2024) remarked, "AI-generated images and responses are true to the dataset, but 'wrong' to the world." As an audience, we might be fooled if we did not know Sydney, but for local citizens who know the topography of that location well, or even more specifically, retailers and office workers based there, this is evidently false.

Viewer Interpretation and Confusion-Neuroscientific Concepts

There is a polarity here. One type of audience has no reason to refute what might lie on Market Street—the image generated is highly plausible as to what "could" be there. Nevertheless, clearly, AI is very wrong in its outcome, and those "in the know" would be irritated that the program had given this return, and most certainly confused. According to neuroscientists, a viewer's cognition and interpretation of what they see, is made via a complex mixture of neural responses:

The impetus for image apprehension comes mainly from the interaction between the brain, body, mind, and environment. Furthermore, many elements in the visual apprehension process activate specific cognitive pathways during this interaction. They contextualize our responses and shape the possibilities of how images are perceived. These elements include semiotic content, emotional valence, arousal, personal intent when viewing, the qualities of the viewing device, the context of use and the source of the image. (Marotta 2023, 33–55)

When we see things, particularly things with which we already have a familiarity, we then base our expectations of a new encounter of similar materials on what is known in neuroscience as "priming." "Priming refers to the process by which exposure to a stimulus influences an individual's subsequent perception, behavior, or response to a related stimulus" (Albornoz and Maestre 2023, 5). Introducing Clarissa Machado's view on at this point is insightful. She was asked if she felt an audience would respond differently to photographs (or at least what "appears" to be photographs) when informed that they are not produced by conventionally understood methodology. Is there a different emotional reaction that reduces their sense of acceptance of "reality" over what they see if it is AI-generated? Her reply references Leon Festinger's (1957) psychological concept introduced in his 1957 A Theory of Cognitive Dissonance, which describes the mental discomfort experienced when we are faced with conflicting information.

In the realm of architectural imagery, this dissonance becomes particularly pronounced when viewers learn that what appears to be a real photograph is, in fact, generated by AI. This revelation often shifts their emotional and cognitive engagement from an initial acceptance to skepticism, profoundly affecting their connection to the image. (Machado 2024)

If an audience has no reason to question an image's provenance via their unintended ignorance, then AI can be believable. Those viewers would then be "hoodwinked" into believing that what they were being shown was true. Moreover, it would not matter—until or unless they became aware that it was not, in fact, a genuine built environment. So, could the software be working from old photographs of Market Street? Might this have been a longdemolished building, subsequently replaced? Yet, this is also highly improbable, after all, Midjourney shows us a junction, and a building straddling two street frontages, which is not the case in Sydney.

For the Bangkok Tea-house, Leverett used increasingly detailed prompts to try and drive the intended outcome. In this next investigation, he was asked to retain the Sydney address and add the single word "fog" to see what the software would produce. A "seed number" is used in Midjourney to generate each unique output (Deng 2024). Retaining the same seed number and building within the prompt should help produce consistencies as well as variations, for example promoting the addition of a curtain wall to an east elevation while leaving the south façade intact as before. Except in this trial, it does not (Figure 14). It fails to maintain the visual integrity of its predecessor. Figure 14 shows that despite the sole addition of the word "fog," a brand-new building has been generated, albeit one shrouded in mist. The time frame for the architectural style has significantly shifted by approximately one hundred years too. Again, this may be confusing, frustrating perhaps for a user wishing to maintain consistency in returns.



Figure 14: Prompt Given to Midjourney—Photograph of 100 Market St, Sydney, NSW, 2000 Australia, fog.—ar 16:9—seed 301—v6. 2024. https://alpha.midjourney.com/jobs/eb50919a-3bcb-4de8-9705-e3026bb2230a?index=3 Source: Damon Leverett

Limitations of Creative Suites for Professional Architectural Documentation

Midjourney (2025) is marketed as a "creative suite." This program works on the principle of "Diffusion," which fosters new interpretations based on the visual datasets on which it is trained. Midjourney learns through a "forward and backward" process of understanding the ingredients that make up the visual references for a subject and then learning to "unpick" them one-by-one, to get back to the original data. By learning the recipe, the program develops the processes necessary to generate new data and most plausibly, creative outputs, but not scientifically accurate reports on a time, place, or subject. Operated using its Discord bot within its own Discord channel, users make an account to run the program and create their images (Altexsoft 2023). Perhaps, given Midjourney's (2025) mantra that invites the user to "Unleash your creativity, [the software] is your first step to crafting stunning images and exploring endless possibilities," it would be unreasonable to ask it to mimic *known* built environments. As Leverett (2024) retorts:

Why would we create an AI photo of a specific place and time when an actual photo is the only medium that can capture all the dimensions of the image that is desired? Even if AI could create an exact place and time, what about the people and other elements in the view?

Nevertheless, while no specific software is cited here, there are more scientifically focused AI programs in the market that may be harnessed to assist with the documentation and preservation of architectural design and histories. Working with controlled image datasets together with the potential to drive and refine what Xu et al. (2024) refer to as "Golden" seeds, users should be able to create image returns far more consistent in style than possible with expansive creative suites such as Midjourney. These would then have a role in targeted application. Architect Abíodún Adéseye (2023) expresses that AI is highly useful in architectural history and conservation work, for architectural style recognition and digital modeling to help predict future issues with structures based on previous visual timelines. These may aid conservation and preservation. There are still challenges to such uses however, caused by a "risk of unintentional revisionist history [which] poses ethical concerns about the accuracy of restored architectural elements." The choice of *which* programs to utilize for such work is key here, a creative suite, while powerful in scope, may bring significant inaccuracies to returns. This necessitates selection of a software for the task that adheres to datasets with specified alignment criteria and to be accompanied by controlled and measured use by its operators.

The Case for Understanding Provenance

So, how an image is produced, and with which type of software, significantly affects the output. In the current case study, the mimicking of architectural styles is shown to be entirely

possible, but viewer ignorance of a picture's provenance will likely become a barrier to certain types of communication and by proxy, "trust" in authenticity. The "authority" standing behind an image will logically contribute to the "trust" of that picture, over and above its actual provenance, lens-based or otherwise.

In other words, if Yale University Press, the publishers of renowned architectural guides were to reproduce Leverett's Figure 14 Midjourney image in a book on historic architecture and stated, "this is accurate," would we be content to accept it, and *not* question it, even if the source was not lens-based? Furthermore, if Yale had not stated the origin, would humans simply believe in it regardless? We trust published drawings of historical structures—e.g., an eighteenth-century line drawing of a Corinthian column capital or the frontage of Bath Cathedral with Jacob's ladder on it as being able to represent what is there. Why not an AI representation of it? This is an example of need, where depicting a typical architectural "style" versus requiring an authentic case study may in fact be preferable to mitigate risk of too much "individuality" in a building type from site-to-site and case-by-case.

The fact that photography, a seemingly dispassionate mechanical procedure, captured both relevant and incidental details of a particular object of observation raised questions regarding its utility for scientific illustration in the nineteenth century. Clarifying drawings or diagrams were sometimes needed to make sense of early scientific photography, which might also be clouded by artifacts from the technical process. The understanding of photographs as both too indiscriminate and too specific was thus a facet of debates in the sciences around the adoption of photography. Scientific drawing traditionally depicts a generalized exemplar of the subject matter at hand, distilled from its most common characteristics. (Wasielewski 2023, 191–210, 198)

So, in this regard, the use of creative programs such as Midjourney, which understands and can recreate typical architectural styles and decorative motifs, could potentially be employed in the same manner as an architectural drawing to exemplify a particular look, or a period of design. On this basis, audiences could plausibly invest their trust in the promise. Machado (2024) is able to elaborate on this premise:

In my practice, the credibility of the source is foundational in shaping our trust and acceptance of architectural representations, whether they are traditional or AI-generated. This is aligned with the source credibility theory in psychology, which suggests that information from a reputable source like Yale University Press is typically more readily accepted due to the perceived authority, expertise, and trustworthiness of the publisher.

HAMILTON KNIGHT: ARCHITECTURE SEEN THROUGH THE EYE OF THE LENS AND AI

Audiences are quickly learning not to take what they see as "photographs" at face value any longer though, and genuine lens-based work is now unfortunately being called into question. Figure 15 is a picture by the Copenhagen-based British photographer Alastair Wiper (2023), which was part of his 2020 book *Unintended Beauty*. The picture shows the largest solar furnace in the world at Odeillo, France.



Figure 15: Odeillo Solar Furnace France. 2012 Source: Wiper 2023

This is a strange photograph. The architecture does not follow form as a familiar vernacular seen in utility buildings, and despite widespread publication of industrial sites by notable practitioners such as Bernd and Hill Becher, Wiper's photograph might easily be confused with a "fantastical" output by an AI-generative program (De Duve 2020). After all, we are very aware that the ability for AI to blend seemingly unconnected objects to reveal surreal and exciting imaging possibilities is where creatives are spending much screen time experimenting and innovating with programs such as Midjourney, Stable Diffusion, and DALL-E. Machado (2024) comments:

Alastair Wiper's works further exemplify how genuinely surreal images can stimulate curiosity and wonder. The photographs in his book, though real, challenge our preconceptions about what architectural photography can represent. Where photographs might be challenged as "too perfect," or that the scene depicted looks wrong or false for some reason but can then be proven to be authentic such as Wiper's Unintended Beauty series, will mean a re-evaluation of expectations and enhances the appreciation for creativity and innovation in capturing real-world architecture. The realization that these unconventional images are authentic, once revealed, deeply influences our understanding of authenticity and the potential of architectural representation.

Moreover, we rely primarily on photography's ability to provide a sense of certainty about existence to support our learning. Despite the subjectivity of each image and the eye behind it, as Heather Dewey-Hagborg (2023, 56) writes:

Photography generally implies a subject that can be represented; it seems to have a certain claim on "reality." This is not an abstract or esoteric consideration, but one with social and political consequences for how we see the world around us, and how we assess the truth claims of images that circulate in our networks.

Conclusion

Ultimately, there remains one main thing that marks the difference between us trusting in an AI-generative image and one made with a camera. Although a photographer can massively influence how a photograph looks through their choice of lens, viewpoint, lighting choice, whether it is monochrome, color, straight out of the camera or heavily re-touched, ultimately, the building they are photographing has a physical form. It is there, for us to see, to touch, to examine from all sides, and to validate. Yes, this exists.

It also means that others may add to that evidence, which declares the building to be genuine as well. They may also photograph it, film it, draw it, describe it with words, or use any number of media to describe something that exists in physical three-dimensional form. Each may vary, but they are based on the same source material, and as Machado (2024) confirms, "with photography you may create different images, you can position yourself in a different way, use different lenses and get different results. But the building is always the same, if you go there and see it in person, it's the same."

There is consistency in what is situated at that site. Moreover, this proof with a camera may be validated by one eye documenting it from a variety of angles inside and out, or multiple photographers with different equipment in differing weathers and light doing this. This may be supplemented by autonomous camera systems in the same vicinity too: CCTV, dashcams of passing vehicles, and so on. Each picture will vary, and the site conditions will change, but the structural form of the building stays consistent throughout. The quality of each output may be inconsistent, but nevertheless cumulatively, they accrue a portfolio of visual trust in what exists as a given location.

Hito Steyerl (2009, 86–91, 89) examines what she terms, "the poor image," and while mainly discussing the circulation of video clips and the annexation of images from their original source and context, she makes a solid point about the validity of numerous individual sources in existence at any given time, stating "Poor images are thus popular images—images that can be made and seen by the many." Such images serve to provide authenticity to the notion that "this is how it looks, it does exist" even if the physical aspect of each form is poor in its quality, with excessive pixelation and reduction in quality owing to repeated compression (as seen in lossy files such as jpegs). Architectural historian Nigel Coates (2012, 15) states:

While permanence should be celebrated as a particularly architectural quality, inevitably, we should be curious about its opposite. The difference between a mere image and a work of art lies partly in its endurance—of existence but also of meaning. In architecture, that endurance is both positive and negative, depending on whether the public buys into it or not.

Currently, as Leverett (2024) shows with his Bangkok Tea-house, 100 Market Street, and for those who have used creative AI software suites, each venture brings a new outcome. Repetition is never guaranteed—the black box makes its "magic" each time, delivering either delight or frustration depending on what it chooses to create from our prompts.

For certain intended uses, this is troubling, there is not enough consistency, and moreover, we are unable to contest the software enough to refine what we wish for on a repeatable basis. This fact means that we cannot rely on AI images, because fundamentally, it does not know the world around us, it only knows its dataset. There is no stability in its outputs, we cannot control them precisely. The author set out to prove that provided AI could be tasked to depict architectural spaces and places, provenance of the image is unlikely to be an issue for many audiences. However, clearly in the case of time-bound accuracy, it cannot, and for those who need trust in the architecture depicted in a certain location, this does not work, and architectural photography prevails, especially when validated by an accumulation of examples.

However, some AI text-to-image outputs are very, very convincing, and suffice as a strong example of type or form, and in that context, it does not matter what the provenance is, when we want ideas and suggestions or typicality. For these instances, AI can be a terrific muse. Indeed, Leverett is one of many thousands of users who find the software to be a creative tool for shaping new possibilities and opportunities to influence the form of architecture, which then may come to pass, as opposed to already existing.

When it comes to making an image of an existing built environment, from the newest of constructions to the oldest of antiquities on UNESCO World Heritage sites, for now at least, nothing is better than a skilled architectural photographer working on site. A trained, creative eye and an editor's toolbox may focus a viewer's gaze, but each audience must understand what they are looking at. They need to know that they may trust it enough to act as their guide to that building without needing to see it for themselves in three dimensions. "Without cognitive effort, without specific needs, without education, our receptors fall asleep, hypnotized by the interfaces and artificial environments we inhabit, which reduce our autonomy, our singularity and our capacity for discernment" (Keller 2021, 6). This issue adds to the growing calls for authorities to implement a clear labeling of provenance, a subject that is already under debate, with mixed outcomes observed in initiatives sled by platforms such as Instagram. This was introduced in spring 2024 and PetaPixel's Matt Growcoot (2024) reported on indignant reactions to the seemingly *ad hoc* labeling of photographs that accurately or otherwise had been tagged as the outcome of AI source or editing.

Kindle Direct Publishing (2025) is another example of policy requiring labeling on image-based works submitted, stating "We require you to inform us of AI-generated content (text, images, or translations) when you publish a new book or make edits to, and republish, an existing book through KDP (Kindle Direct Publishing) AI-generated images include cover and interior images and artwork." We are now used to academic bodies, publishers, and universities requiring labeling of works and statements of provenance for written submissions, and gradually these policies will need to encompass all types of media in order to mitigate audience confusion and instead provide accurate cognition.

This research concludes that for complete interpretation and understanding of real-world architectural environments, we do need to be sure that the images we are viewing relate conclusively to the built form they represent. Again, context is everything: if the institution behind the image represents "trust and authority" then readers will have faith in what they see. It is down to the brand. Therefore, if it is the product of a text-to-image program, and Yale University Press say, "yes, this is correct," then readers will not question it. The same AI image in a different context or use, *or* exampled by a body unknown to the viewer, *or* a viewer without a degree of visual literacy *or* from a source not regarded by that viewer as "trusted" may not receive the same authority. We are learning to navigate. Our education is vital. Labeling is vital. Then, and only then, it will cease to matter whether it was created by a camera and a human, or a series of prompts and a machine.

Acknowledgment

This ethically approved research has been carried out with gratitude to Damon Leverett and Clarissa Machado for their time and contributions. Both sets of inputs have exceeded the scope of conventional research participation. Information about Damon Leverett may be found here: https://www.linkedin.com/in/damonleverett/. For information about Clarissa Machado's research: https://www.movingboundaries.org/clarissarodriguesmachado.

Written questions and responses from both interviews discussed in this output, together with full prompt text for twenty, hundred and thousand-word Chat-GPT prompts for use by Midjourney Version 6 by Damon Leverett (2024) are stored at Zenodo: https://doi.org/10.5281/zenodo.14926151 and are available for inspection by request to the author.

AI Acknowledgment

The author declares that generative AI or AI-assisted technologies were not used in any way to prepare, write, or complete this manuscript. The author confirms that they are the sole author of this article and takes full responsibility for the content therein, as outlined in COPE recommendations.

Informed Consent

The author has obtained informed consent from all participants.

Conflict of Interest

The author declares that there is no conflict of interest.

REFERENCES

- Adéseye, Abíodún. 2023. "Rediscovering the Past: Recognition and Reconstruction of Architectural Historical Design Elements with Artificial Intelligence." *Medium*, November 12. https://medium.com/@abiodun.o.adeseye/rediscovering-the-pastrecognition-and-reconstruction-of-architectural-historical-design-elementsb893f461884c.
- Adobe. 2025. "Get Professional-Quality Results with Generative Fill." https://www.adobe.com/uk/products/photoshop/generative-fill.html.
- Albornoz, Yohany, and Gladys E. Maestre. 2023. "Insights from a Study on Multimodal Experiences in a Plaza." *Journal of Eco+urbanism and Neuroarchitecture (JEN)* 1 (1). https://scholarworks.utrgv.edu/som_pub/1098/.
- Altextsoft. 2023. "AI Image Generation Explained: Techniques, Applications, and Limitations." July 10, https://www.altexsoft.com/blog/ai-image-generation/.
- Barthes, Roland. 1977. "The Rhetoric of the Image." In *Image, Music, Text.* Translated by Stephen Heath. Hill and Wang.
- Beaumont, Romain. 2022. "LAION-5B: A New Era of Open Large-Scale Multi-Modal Datasets." March 31, https://laion.ai/blog/laion-5b/.
- Burgin, Victor. 1987. Thinking Photography. 4th ed. Macmillan Education Ltd.
- Coates, Nigel. 2012. Narrative Architecture. Wiley.
- Cruz, Edgar Gomez, and Eric T. Meyer. 2012. "Creation and Control in the Photographic Process: iPhones and the Emerging Fifth Element of Photography." *Photographies 5* (2): 203–221. https://doi.org/10.1080/17540763.2012.702123.

De Duve, Thierry. 2020. Bernd & Hilla Becher: Basic Forms. Prestel.

- Deng, Edward. 2024. "Midjourney Seed Tutorial—How to Use Seed in Midjourney for Beginners." AI Chronicler, February 29. https://aichronicler.com/how-to-use-seed-inmidjourney/.
- Dewey-Hagborg, Heather. 2023. "Generative Representation." In *A Cat, A Dog, A Microwave—Cultural Practices and Politics of Image Datasets*, edited by Nicolas Malevé and Ioanna Zouli. The Photographers Gallery.
- Elcott, Noam M., and Tim Trombley. 2025. "How Can Image-Makers Open Up AI's Mysterious 'Black Box'?" *Aperture*, March 7. https://aperture.org/editorial/how-canimage-makers-open-up-ais-mysterious-black-box/.
- Festinger, Leon. 1957. A Theory of Cognitive Dissonance. 1st ed. Stanford University Press.
- Growcoot, Matt. 2024. "Instagram Photos are Being Labelled 'Made with AI' When They're Not." *PetaPixel*, May 28. https://petapixel.com/2024/05/28/instagram-photos-are-being-labeled-made-with-ai-when-theyre-not/.
- Hacking, Juliet. 2012. Photography, The Whole Story. Thames & Hudson.
- Hamilton Knight, Martine. 2022. "The Pevsner Way of Seeing, A Paper on Contriving 'Timelessness' in The Buildings of England series of Books." *Quart* 1 (63): 82–101. https://doi.org/10.11588/quart.2022.1.90758.
- Hamilton Knight, Martine. 2023. "New Topographics: Man-Altered Landscapes Become Machine-Altered Landscapes." Presented at the Photography and Artificial Intelligence Conference, Royal Photographic Society, Bristol, October 9–10, 2023. https://rps.org/about/artificial-intelligence-and-the-rps/conference-outline.
- Hartwell, Clare, Elizabeth Williamson, and Nikolaus Pevsner. 2020. Buildings of England: Nottinghamshire. Yale University Press.
- Kindle Direct Publishing. 2025. Content Guidelines: Artificial intelligence (AI) Content (Text, Images, or Translations). https://kdp.amazon.com/en_US/help/topic/G200672390.
- Keller, Milo. 2021. "Teaching the Machines, Don't Panic, Learn How it Works." In *Automated Photography*, edited by Milo Keller, Claus Gunti, and Florian Amoser. Morel Books.
- Leach, Neil. 2022. Architecture in the Age of Artificial Intelligence: An Introduction to AI for Architects. Bloomsbury.
- Lei, Celina. 2024. "AI Fomo? Why Cool Heads are Needed in the Arts." *Arts Hub*, February 16. https://www.artshub.com.au/news/features/ai-fomo-why-cool-heads-are-needed-in-the-arts-2614040/#:~:text=Different to the fear that, left behind in the abyss.
- Leverett, Damon. 2024. "Interview by Martine Hamilton Knight." July 7. https://doi.org/10.5281/zenodo.14926151.
- Lutkevitch, Ben. 2023. "What is AI Alignment?" *TechTarget*, May. https://www.techtarget .com/whatis/definition/AI-alignment.
- Machado, Clarissa. 2024. "Interview by Martine Hamilton Knight." July 27. https://doi.org/10.5281/zenodo.14926151

- Marotta, Thomas. 2023. "Feeling, Thinking, and Not Seeing: How Images Engage and Disengage in an Information-Saturated World—A Neurophenomenological Perspective." *Media Practice and Education* 25 (1): 35–55. https://doi.org/10.1080/25741136.2023.2243376.
- Meyer, Eric T. 2005. "Socio-Technical Perspectives on Digital Photography in Professional Practice." Doctoral Qualifying Paper. Indiana University.
- Midjourney. 2024. "Version 6." https://docs.midjourney.com/hc/en-us/articles /32199405667853-Version.
- Midjourney. 2025. "Getting Started Guide." https://docs.midjourney.com/hc/enus/articles/33329261836941-Getting-Started-Guide.
- Mirzoeff, Nicholas. 2015. How to See the World. Pelican Books.
- OpenAI. n.d. ChatGPT [Large Language Model]. Accessed June 2025. https://chatgpt.com.
- Ploennigs, Joern, and Markus Berger. 2023. "Generative AI and the History of Architecture." *arXiv Preprint*. https://doi.org/10.48550/arXiv.2312.15106. Cornell University.
- Smith, Joel. 2011. The Life and Death of Buildings. Princeton University Art Museum.
- Steyerl, Hito. 2009. "In Defense of the Poor Image." *E-flux Journal* 10. https://www.e-flux.com/journal/10/.
- Van Gelder, Hilde. 2009. "Photography as a Locus for Destabilising Fundamentals: The Multimedia Installations of Philippe Van Snick." In *Take Place—Photography and Place from Multiple Perspectives*, edited by Helen Westgeest. Valiz.
- Wakeling, Christopher, and Nikolaus Pevsner. 2024. *Buildings of England: Staffordshire*. Yale University Press.
- Wasielewski, Amanda. 2023. "Authenticity and the Poor Image in the Age of Deep Learning." *Photographies: The Politics and Practices of Computational Seeing* 16 (2): 191–210. https://doi.org/10.1080/17540763.2023.2189158.
- Wiper, Alastair. 2023. Unintended Beauty. 2nd ed. Hatje Cantz Verlag.
- Xu, Katherine, Lingzhi Zhang, and Jianbo Shi. 2024. "Good Seed Makes a Good Crop: Discovering Secret Seeds in Text-to-Image Diffusion Models." In arXiv Preprint arXiv:2405.14828v1. May 23. https://arxiv.org/pdf/2405.14828v1.
- Yale University Press. 2024. Pevsner Architectural Guides. Yale University Press.

ABOUT THE AUTHOR

Martine Hamilton Knight, DLitt (Hon), FHEA, FRPS: Senior Lecturer, Photography, Design & Digital Arts, Nottingham Trent University, UK Email: martine.hamiltonknight@ntu.ac.uk