THE GLASGOW SCHOOL OF ART

The Investigation of the Capabilities of the Black Box and Grey Box Creative AI Models in the Architectural Design Process

Mstislav Kochkin

Mackintosh School of Architecture

The Glagow School of Art

This thesis is submitted for the degree of

Master of Architecture (by Conversion)

January 2024

Abstract

Over the past decade, the digital transformation of architecture amid the reportedly happening 4th Industrial Revolution (4IR) allowed Artificial Intelligence (AI) to start gaining its niche in architectural design. AI-powered tools are being introduced to BIM software, such as Revit, and top-tier practices, such as Zaha Hadid Architects or Coop Himmelb(l)au try using Midjourney for architectural visualisations, and developing in-house neural networks (NNs), while governing bodies such as RIBA are working towards its legislation and rational regulation.

Bernstein and Chaillou argue, that AI is the next stage of architectural technological evolution after CAD and BIM. The AI-powered optimisation tools have already taken a firm position due to reliance on rules and quantifiable parameters, that match with the trend towards design process and supply chain automation and rationalisation. However, the use of tools like ChatGPT, Midjourney or Stable Diffusion (SD) in the "creative" domain of architecture has caused discussions about their impact. Now "creative" tools are falling behind due to AI's inability to extrapolate, difficult translation of the visual language into text, frequent misunderstanding of the metaphorical meanings, undefined assessment criteria for AI-generated output, and users' mistrust of available tools due to their "black box" nature.

This research aims to assess the potency of the "creative" tools and discuss their capability to augment architectural thinking and design processes. Using the Final Design Thesis (FDT) as the study and training dataset subject, I tested four "black box" and two "grey box" models in a two-way process. Firstly, existing architectural representations were translated into semantic maps and fed to Dall-E, Midjourney, SD and LookX to determine which types of mediums they could generate. The FDT was also used as a dataset to train SD and LookX to determine whether AI can create an architect's work without the loss of accuracy and disclose its latent potential. Secondly, the same AI tools + ChatGPT were tested to assist in developing FDT from scratch to determine their capabilities in research and development of the original design solutions, and at which design stages devoiding of optimisation tools is impossible. The research outcomes should demonstrate the advantages and limitations of existing "creative" AI tools, and possible improvements in the architect's design process, and give an understanding of some ethical implications behind this.

Due to the topic's novelty, much research and literature that is yet to be published can contain fundamentally new information, which might change the topic's agenda. Therefore, the terminology and research outcomes are fully based on the available sources and all the judgements and speculations in this research derive from them.

Declaration

I confirm that this work is my own, undertaken for the degree of Master of Architecture by Conversion, and that it has not been submitted for any other award or assessment.

Signed: [redacted] *Date:* ***15 January 2024***

Table of Contents

Word Count: **10,613**

Glossary

4th Industrial Revolution (4IR) – The next phase in the digitalization of the manufacturing sector, driven by disruptive trends including the rise of data and connectivity, analytics, human-machine interaction and improvement in robotics. (McKinsey & Company, 2022).

Artificial Intelligence (AI) – 1. A technology, which mimics human cognitive functions. 2. Digital tools, which utilise programs, that mimic human cognition to successfully commit the tasks (Del Campo, 2022a, p. 17).

Artificial Neural Network (ANN) – Machine learning systems, which contain an input layer, one or more hidden layers, and an output layer within their structure. They are inspired by the human brain, mimicking the way that biological neurons signal to one another (IBM, 2023a). The abbreviation ***NN*** is also applicable in this work.

Central Processing Unit (CPU) – A complex set of electronic circuitries, which is the primary component of a computer and interprets, processes and executes instructions from the operating system and apps (ARM, 2023a).

Beta Testing – A final round of testing before the release to a wide audience, at which the developers collect feedback provided by the users, who test the program. Beta testing can be ***closed*** (limited to a specific group of testers) and ***open*** (available to all users to test).

Checkpoint – Pre-trained Stable Diffusion model, which is used for generating images of a particular object or style. The term derives from a .ckpt file format, in which pre-trained models are often saved.

Cloud-Based Services – Online applications, in which third-party providers allow the customers to leverage powerful computing resources without having to purchase or maintain personal hardware and software (Hewlett Packard Enterprise, 2024).

Constraints – In parametric modelling: values, combinations of values or relations between parameters using which will result in unacceptable solutions (Veloso and Krishnamurti, 2021, p. 40). In a wider context: any architectural solutions, which AI must avoid to meet the architect's input task (Barker, 2023).

Contextualisation – The discussion of the relationships between buildings and their environment (Del Campo, 2022a, p. 33).

ControlNET – A model for controlling image generation by diffusers by conditioning them with an additional input image. Conditioning input types, which include edges, surface normals, depth, etc. are analysed by the sys-tem and used as a template for the AI to generate new images (Hugging Face, 2023a).

«Creative» AI Tools – In this work: AI-powered models, which allow for output generation with perceptive skills, visual or linguistic input and implicit knowledge, rather than numeric parameters and explicit logic, and can be used in the early stages of design ideation process.

Deep Neural Network (DNN) – A neural network that consists of more than three layers (IBM, 2023a).

Digital Transformation – A fundamental shift in architectural culture facilitated by the adoption of innovative digital technologies (Microsoft and RIBA, 2019, p. 11).

DreamBooth – A fine-tuning technique that updates the entire diffuser by training on few images of an object or style. It works by associating a special word in the prompt with the example images (Hugging Face, 2023b).

Feature Identification – Identification and recognition of particular architectural elements and properties, which allow to compare buildings between each other and categorise them. (Del Campo, 2022a, p. 34-34).

Functional Requirements – The general notion of specific functionality that the system or object must be able to perform.

Genetic Algorithm – A randomised search algorithm that imitates C. Darwin's theory of natural selection. It uses selection, crossover and mutation operations to create a set of randomly generated solutions, which are then assessed against fitness to parameters. The worst fitting options are eliminated with each generation step, till few highly fit options remain (Chiroma et al., 2017, pp. 1544-1545).

Generative Design – A design development process, in which the designer inputs design goals into the model, along with parameters and constraints. Using the power of computation the model automatically explores all the possible permutations of a solution, quickly generating design alternatives (Autodesk University, 2023).

Generative Model (GM) – A design synthesis model, which utilises various forms of decision-making and operates in two stages. In the formulation stage, the generator's algorithm is encoded by the designer, and the input and output representations are developed. In the execution stage, the designer interacts with the generator, changing the input and observing design alternatives for feedback (Veloso and Krishnamurti, 2021, p. 30).

Grammar Approach – In linguistics: the rules of formation, structure, and assemblage of language. It studies and formulates the rules (also called heuristics) by which words come together to create phrases and their function in the language, rather than their meaning. In architecture: organising building forms based on their function (Chaillou, 2023)

Graphics Processor Unit (GPU) – An electronic circuit designed to process images and accelerate the rendering of 3D computer graphics on computers, smartphones, and gaming consoles (ARM, 2023b).

Graphical User Interface (GUI) – The interface that allows users to interact with electronic devices through graphical elements, such as tabs, buttons, scroll bars, menus, icons, pointers and windows, instead of text-based commands (Indeed, 2023).

Immediate Context – A specific site in a certain place and particular culture (Chaillou, 2022, p.

197).

Low-Rank Adaptation (LoRA) – A fine-tuning technique that injects trainable layers into the structure of a pre-trained model (Cuenca and Paul, 2023).

Machine Learning (ML) – Use of algorithms, trained to make classifications or predictions, based on input, data, and experience (IBM, 2023b).

Metaheuristics – A set of optimization techniques that find a set of overall best solutions for a given problem by iteratively sampling solutions and using performance criteria to generate better and better outcomes (Autodesk University, 2022).

Metaoptimisation – Tuning parameters of an optimisation algorithm with use of a higher-level optimiser (Opara and Arabas, 2012, p. 110).

Narrow AI – AI, which is designed to excel in one particular task in the same way a human, but may struggle to commit tasks outside its expertise (Nancholas, 2023).

Neural Style Transfer – An optimisation technique, in which a content image and a style reference image are blended so the output image looks like the content image, but represented in the style of the style reference image (TensorFlow, 2024).

Noise – In diffusers: ***Gaussian noise***

Non-functional Requirements – The general notion for characteristics, qualities or attributes which the system or object must have to meet user expectations.

Phenomenon – In this work: specific data observed by the ML program to capture its characteristics for future prediction and mimicing under new parameters (Chaillou, 2022, p. 65).

Plug-in – A small computer program, which makes a larger one work faster or have more functions and features (Cambridge Dictionary, 2024).

Random-access memory (RAM) – The computer's short-term memory, which stores the data, necessary for the processor to run applications and open files (Intel Corporation, 2023).

Seed – A numeric parameter, which instructs the NN on which image to generate from random noise. If the same seed number and prompt are used, similar images will be generated (Midjourney Inc., 2024).

Stable Release – A version of a program, which has passed all the testing and validation stages and is safe for use by wide public.

Semantic Approach – In linguistics: formulation of sentences using associations, rules and constructs through the lens of signification and their meaning in the language, rather than their function. In architecture: organising building forms based on their meaning and significations (Chaillou, 2023).

Visual Programming – A computer programming method, in which the user connects components

to define the relationships and the sequences of actions that compose custom algorithms. Algorithms can be used for a wide array of applications, from processing data to generating geometry in real time and without writ-ing text code (Dynamo, 2024).

Introduction

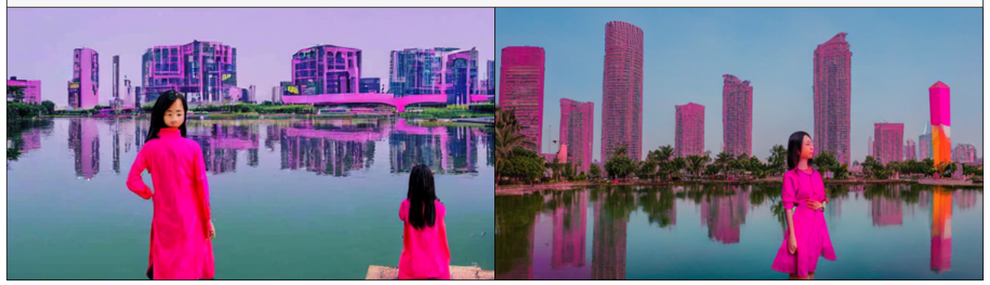
AI has been an integral part of our lives since its inception in the 1950s and gained its niche in our routine amid society's rapid digitalisation. We rely on digital devices, such as phones or computers and do not realise that many previously difficult or impossible tasks are now solved with AI. AI forms our credit score, and music and film recommendation lists, filters mailbox spam, finds potential friends on social media or plays chess against us (Leach, 2022b, pp. 1, 8). The invention of deep learning (DL) in the 2010s introduced AI in the creative industry. AI-generated music is now available on streaming platforms. TextToText chatbots write poems. Publicly available TextToImage tools create evocative artworks, which quickly evolved from "abstract mess" to highly detailed state-of-the-art pieces ***(Table 1.1)***. In 2018, an AI-generated painting was sold for $432,500 (Leach, 2022b, p. 63). However, such rapid development sparked fierce discussions in the creative community. Some question AI's "black box" nature and demand strict regulation or even a ban due to possible copyright infringement and job deprivation for young employees (Combrinck, 2023, pp. 22-26). Others believe that it can become a valuable assistant in speeding up the workflow and pushing output to the next level (Bradley, 2022, p. 69).

Architects' aspirations to design better buildings and places and improve living standards motivate them to learn new technologies, which makes AI's adoption in architecture inevitable. The report by RIBA and Microsoft demonstrates that in 2018 5% of UK architectural firms were already using AI and machine learning (ML), while 19% planned to use it within the following 5 years ***(Figure 1.1)*** (Microsoft and RIBA, 2019, pp. 11, 20). The survey results of ten architects and researchers from world-famous architectural bureaus, presented at the 2020 S.Arch Conference, predicted ML will become the most useful computational technology within the following decade ***(Table 1.2)*** (Gallo, Wirz and Tuzzolino, 2020, p. 200).

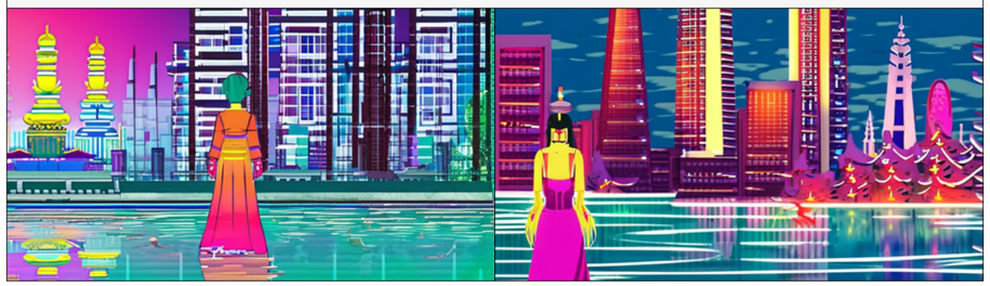
Table 1.1: The Evolution of AI-Generated Art

Prompt: *Girl standing at the pond in the futuristic oriental city with pagoda, and skyscrapers in the background, neon colours, red and purple colour scheme*

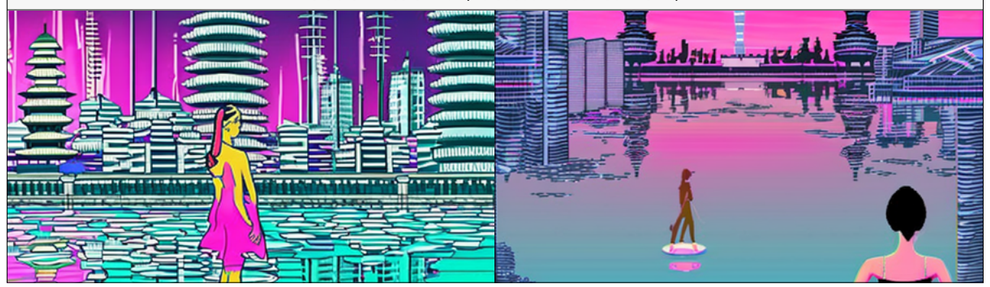
Stable Diffusion 1.5 *(Released: Octoder 2022)*



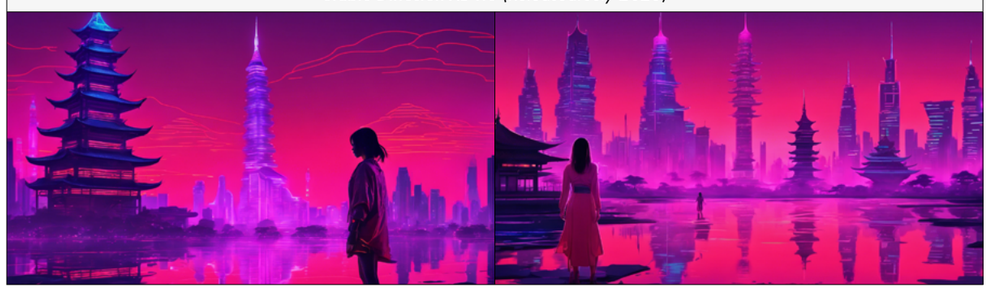
Stable Diffusion 2.0 *(Released: November 2022)*



Stable Diffusion 2.1 *(Released: December 2022)*



Stable Diffusion XL 1.0 *(Released: July 2023)*



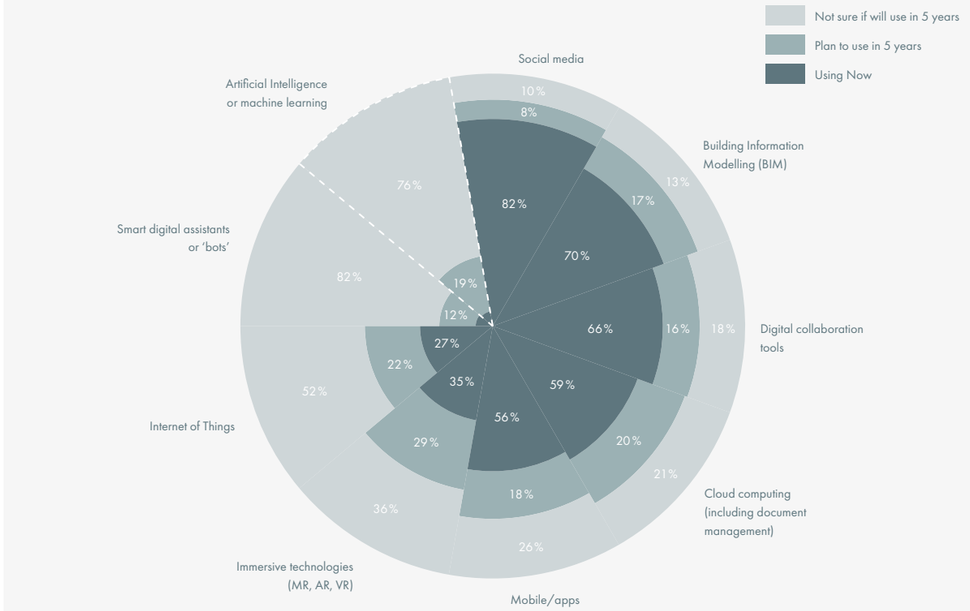


Figure 1.1: Digital technologies used by the UK practices in 2018 *(Microsoft and RIBA, 2018, p. 20)*

Table 1.2: The 2020 survey on the most useful technologies in the following 10 years
(Gallo, Wirz and Tuzzolino, 2020, p. 200).

Interviewee	*BIM*	*IOT*	*AR*	*VR*	*ML*	*Digital Manufacturing*	*Other*
Arthur Mamou-Mani (Studio Momou-Mani)	5	2	3	1	4	7	6
Al Fisher (Buro Happold)	6	2	4	1	3	5	7
Andreas Klok Pedersen (BIG)	7	2	4	3	6	5	1
Aurelie de Boissieu (Grimshaw)	1	5	4	3	7	6	2
Daniel Davis (WeWork)	7	5	4	3	2	1	6
Edoardo Tibuzzi (AKTII)	2	5	3	3	4	6	7
Harry Ibbs (Gensler)	1	2	5	4	7	6	3
Irene Gallou (Foster+Partners)	2	5	4	4	7	6	3
Pablo Zamorano (Heatherwick Studio)	3	4	5	6	6	5	7
Xavier de Kestelier (Hassell)	3	6	1	2	7	4	4
Total	***37***	***38***	***37***	***30***	*53*	***51***	***46***

Since then, AI has been expanding in the industry quicker than expected, thanks to the keen interest of recognised visionaries on one hand, and the profession's strong digitalisation on the other. Zaha Hadid Architects' Patrik Schumacher believes that AI will improve architects' creativity, productivity, and judgement skills ***(Figure 1.2)*** (Schumacher, 2022). Coop Himme(l)blau is developing an in-house NN, which could disclose the latent creative potential of their works ***(Figure 1.3)*** (dPrix *et al.*, 2022, p. 21). In 2020, Autodesk acquired the AI-powered Spacemaker (now Autodesk Forma) for $240 million (Bernstein, 2022, p. 8).

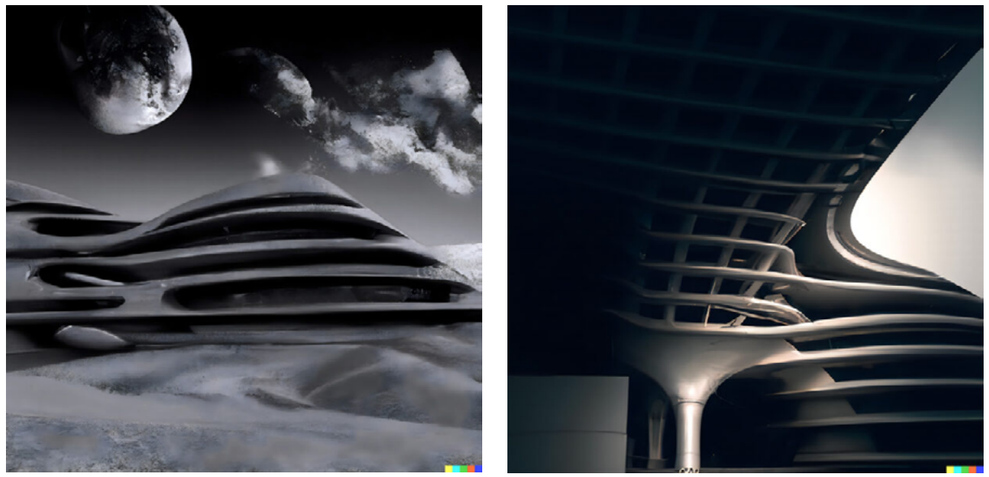


Figure 1.2: AI-generated images by Zaha Hadid Architects *(Zaha Hadid Architects, 2022)*

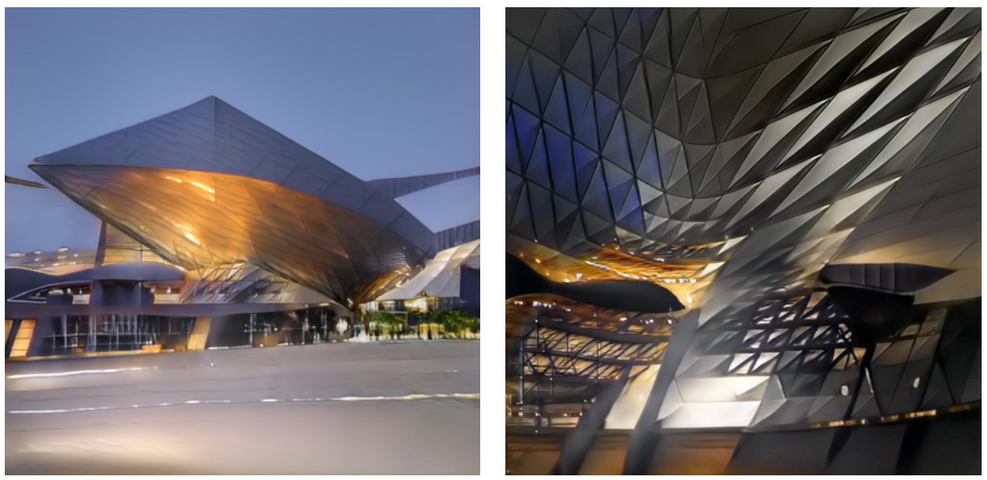


Figure 1.3: AI-generated images by Coop Himmelb(l)au *(Coop Himmelb(l)au, 2022)*

The iterative and complex architectural process fuses engineering formalism with artistic abstract approaches. Architects, apart from developing evocative design solutions, using various media (traditional and CAD drawings, physical and 3D models, etc.) must study urban design, building methods and technologies, environmental systems, architectural economy, law and ethics, the history of art and architecture (Bernstein, 2022, pp. 24-25). This divergence resulted in the proposal of two schools of thinking about AI in architecture: optimisation and creativity. Optimisation explores the possibilities for working based on quantifiable parameters, while creativity is driven by intuition and sensibility, which are hard to translate into code (Del Campo and Leach, 2022, p. 7).

This work aims to investigate the creative potential of AI in architecture at its current stage of development in two parts. The theoretic part will explore AI's principles of work and place in the evolution of architectural digital toolset and discover the optimisation and creative aspects of the existing AI tools. In the practical part, using the 5th Year's FDT as the subject of study, the capabilities of different AI tools will be tested within its design process. Attempting to recreate the FDT with known final output and as if it was developed from scratch, the investigation results will allow an understanding of the nature of AI creativity and which new design approaches it offers, what types of architectural solutions and representations it can produce, and what are the existing practical obstacles and ethical implications for its use. The results will help us understand the possible AI's impact on architectural professional culture and how individual architects, architectural firms, and clients might have to adapt to this new reality.

Literature Review

AI's fast-paced development, multiple related unresolved issues, the absence of an agreed public and professional opinion, and regulations still being in the discussion stage require maximum care in comprehension and interpretation. Additionally, the topic's multifaceted nature demanded an extensive search for scientific work to study AI operation principles, acquire extra knowledge about computer and data science, and improve skills in the CAD toolset. Unthorough research can not only result in formulating a weak argument and methodology but also in biased, ambiguous and controversial judgements.

The knowledge about the existing AI types, their architecture, functions and areas of use was gained from published journal articles and technical papers by independent researchers. The research, conducted by the employees of software and hardware development firms (e.g. Nvidia, Autodesk, IBM, Microsoft and OpenAI) were found on their official websites and verified blogs. Other works (e.g. research papers on DreamBooth and LoRA) are yet to be published and are often presented at AI-themed conferences and are shared via specialised public resources, such as GitHub, Hugging Face and arXiv, alongside the codes and instructions for software installation and use. However, the knowledge and accessibility gap between architecture and computer science complicates studying these sources because explicit programming and maths are not in the architectural curriculum. Written by trained programmers and scientists for their colleagues, the publications often contain many terms, complex mathematical equations and formulas, and coding algorithms, which require specialised knowledge to understand. The official manuals and glossaries by software developers and articles by industry-specialised journalists became possible solutions. They use more generic language to cover a broader audience and provide less in-depth, but still sufficient explanations, allowing readers without IT training to better understand scientific sources and make judgements. They also provided more comprehensive software installation guidance and usage tutorials, which speeded up the investigation process and prevented downloading software incompatible with used hardware and unsuitable for the investigation objectives.

The study of AI in architecture primarily relied on journal and online articles and published books by architectural theorists and practitioners. N. Leach, M. del Campo and S. Chaillou review the history, evolution and current achievements of AI in an architectural context, share their research finding and own projects, explain which AI tools are already exploited in the profession, and which ones have the potential to be adopted in future and refer to industry specialists. They try speculating about how AI can change architectural workflow and synthesis, and discuss the existing economic, political and ethical implications behind it (Chaillou, 2022, pp. 9-10; Del Campo, 2022a, pp. 17-19; Leach, 2022b, pp. xvi-xvii). The monograph 'Routledge Companion to Artificial Intelligence in Architecture' reviews AI in architecture from four perspectives: theoretical and historical background; tools, methods and technology; use in architectural research; and real-world case studies (As and

Basu, 2021, p. xviii). Architectural practitioners, such as P. Schumacher, T. Fu or W. dPrix deliver their vision of AI via interviews, conference lectures, own websites, and articles in architectural magazines (e.g. Dezeen, Archdaily) and journals (e.g. Architectural Design). They share their AI-based design methods and give advice, raise awareness about the existing public misconceptions, discuss the pros, cons and ethical implications of using AI, and speculate about its future. Despite regulatory institutions such as RIBA admitting in their reports, that AI is inevitable in a digitalising profession (RIBA and Microsoft, 2019, pp. 11, 20), such rapid development makes them concerned about its regulation and efficient embedding to the supply chain. Phil Bernstein's book 'Machine Learning: Architecture in the Age of Artificial Intelligence' published by RIBA investigates which AI strategies are most suitable for different design stages and speculates how they can augment the existing workflow and how architects should change their competency to remain demanded in the new environment (Bernstein, 2022, p. v).

Due to the novelty of the topic, some printed sources, which can contain fundamentally new information, may not be released yet, while works which were published within the last year, might already be outdated. For example, 'Artificial Intelligent Architecture: New Paradigms in Architectural Practice and Production' by F. Jacobus and B. M. Kelly was unpublished at the time of the investigation ***(Figure 2.1)***. Therefore, this study refers to the topic's current stage and is derived from only the existing and publicly available resources. The same applies to AI software. StyleGAN still has no graphical user interface (GUI), requiring explicit coding skills, which complicates the exploitation. The stable release of Dall-E, SD and LookX happened only in 2023. While OpenAI provided the interface for the pre-release versions of Dall-E, Stability AI didn't, which forced users to develop their own GUI (e.g. Stable Diffusion Web UI by AUTOMATIC1111), which may provide an opportunity to use the tool without programming knowledge and offer even more functions than its proprietary counterpart. In the case of unavailable GUI from firms, a user-developed interface could be applied only if it matched the legal norms and was based on the official open-source code. However, unlike a firm, the individual programmer might not consider all the implications, such as protection against malware, unbiased tuning and censoring of sensitive content, so these risks must be thoroughly assessed when using such software (OpenAI, 2022). Midjourney and some

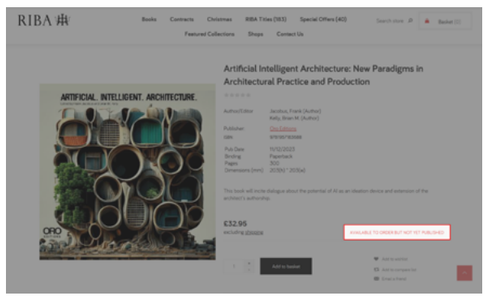


Figure 2.1: The unavailable resources *(RIBA Books, 2023; Infrared City, 2023)*

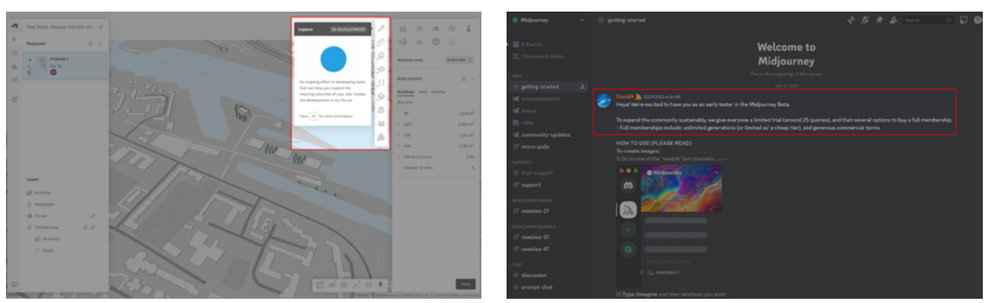
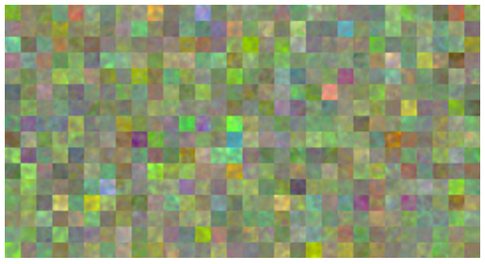


Figure 2.2: Resources in the development stage *(Autodesk Inc., 2023; Holz, 2022)*



```
torch.cuda.OutOfMemoryError: CUDA out of memory. Tried to allocate 1024.00 MiB. GPU 0 has
a total capacty of 8.00 GiB of which 0 bytes is free. Of the allocated memory 20.52 GiB
is allocated by PyTorch, and 1.14 GiB is reserved by PyTorch but unallocated. If reserved
but unallocated memory is large try setting max_split_size_mb to avoid fragmentation.
See documentation for Memory Management and PYTORCH_CUDA_ALLOC_CONF
```

Figure 2.3: StyleGAN crashing during the training attempt.
Top: Fake Samples 1 & Fake Samples 2; Bottom: Memory Error Message

Autodesk Forma functions are in the open-beta stage ***(Figure 2.2)***, which means that they are still in development despite being publicly released and a further improvement in their operation might occur. Others, like InfraRed, are still in the closed-beta stage, making them publicly unavailable ***(Figure 2.1)***. Queuing to subscribe to proprietary software and accurate prompt formulation to avoid referencing existing architects' and artists' works to respect their intellectual property rights are among additional limitations.

Despite the powerful hardware (32GB RAM; CPU: Intel Core i9; GPU: 8GB Nvidia GeForce RTX 2080 Ti), my computer did not meet the minimum requirements for training some AI models, particularly StyleGAN ***(Figure 2.3)***, which required at least 12GB GPU and days or weeks to complete the training (NVIDIA Research Projects, 2023). Such powerful computers were also unavailable on the GSA campus. I could not study much about learning 3D shapes either, since they require equally high computer specifications, and most architectural researchers use cloud-based services to test them (Chaillou, 2022, pp. 111-112). Given that and the short length of the course, it was decided to only focus on generating 2D output with TextToImage tools and some 3D with Autodesk Forma and use less demanding training methods, which will be explained in the methodology.