Architecture

Andri Gerber,
Oya Atalay Franck,
Michael Mieskes (eds.)

Architectural Intelligence in the Age of Artificial Intelligence

[transcript]

Learning from Images

Andri Gerber in Conversation with Philipp Schaerer, Zurich, February 5, 2025

Andri Gerber: We had the pleasure of inviting you to give a talk at ZHAW two years ago, and I was honestly a little surprised when you spoke so positively about AI and its possibilities. It was a time when there was a very negative attitude towards AI, especially in the architectural context, and when I think of your work, I associate it with a very elaborate "digital craft" that is now being greatly shortened by AI.

Philipp Schaerer: Perhaps I should first clarify: when I talk about AI today, I am primarily referring to AI image generation—just as I did back then in my presentation.

Automated image description through captions has already been around for some time and is still widely used today for image tagging. This almost inevitably led to the question of whether this process could be reversed—i.e., whether images could be generated automatically from text fragments. A first significant step in this direction was taken in the mid-2010s by Elman Mansimov and his team. They developed an AI image-generation prototype that showed that this was possible in principle. This opened the door, so to speak, for the further development of more powerful models, which were refined further and further in the following years and finally made available to the public.

From 2020 onwards, AI-generated images spread rapidly—especially via social media. I think we were all fascinated by these images: their unusual softness, their precise attention to detail, and their deceptively real, almost photographic aesthetic. The countless curtains were particularly striking in these architectural fantasies [laughs].

This realistic aesthetic was simply astounding. Suddenly, anyone could translate their architectural fantasies into photorealistic images—without much effort, using text instructions alone, without any in-depth understanding of images, and without any knowledge of the optical laws of photography.

I found this fascinating, but at the same time it raises fundamental questions about authenticity. The fragile boundary between appearance and reality, between truth and staging, has preoccupied me since my first series of images, "BILDBAUTEN" (2007)—and has become all the more relevant with these new technologies.

Fig. 13: Philipp Schaerer, Bildbau No 1, 2007

AG: How do you deal with this new condition in your work?

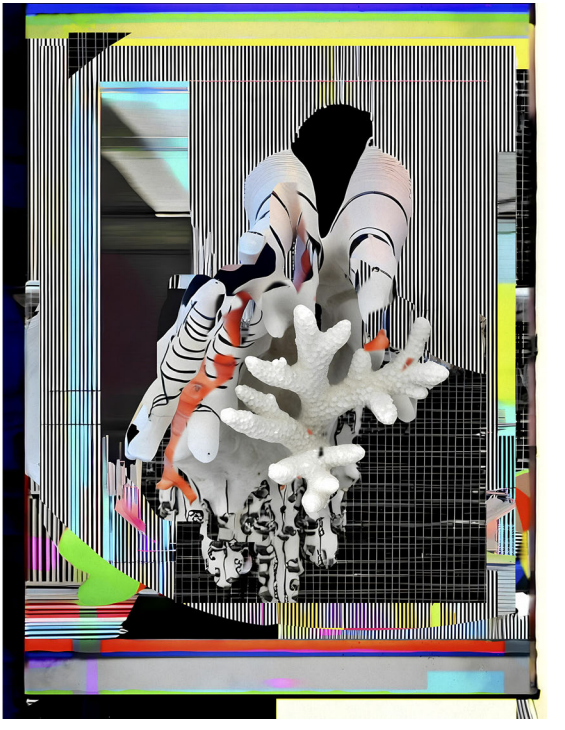
PS: The effort required to generate such images is relatively low, and the resulting image often only approximates the idea you originally had in your head.

That's why it's out of the question for me to consider an AI-generated image as the end product.

In my works, AI-generated images serve mainly as building blocks—be it in the form of textures that I have specifically generated or as representational fragments that I then integrate into a pictorial ensemble using a conventional image montage. One example of this is my latest work, "Crossbreeds—Imaginary Still Lifes," in which individual fragments come from AI-generated images or were rendered directly into the image by me.

For me, making pictures is still a manual process, a continuous development and creation in which I always want to consciously decide in which direction the picture should develop.

Fig. 14: Philipp Schaerer, from the series "Crossbreeds – Imaginary Still Lifes", 2024



AG: You mentioned the subject of "authenticity." If you look at your earlier works—I'm thinking in particular of your "BILDBAUTEN"—the observer always asked himself whether these architectures were real or not. The pictures were very ambivalent, which was also their great quality.

PS: Technically speaking, the "BILDBAUTEN" series are image montages, i.e., composites of various photographic fragments—image constructions that have been created "by hand." Despite minor image inconsistencies or irritations, they deliberately play with photography's claim to credibility.

In the context of AI image generation, it is important to emphasize that such image montages previously required in-depth prior knowledge and technical expertise. The seamless and deceptively realistic weaving of image fragments into an image surface required experience in image processing—just as the use of 3D-rendering programs to simulate photographic representations demanded a certain amount of specialist knowledge.

However, this has changed fundamentally with the advent of AI image generators. Today, simple text-based descriptions (prompts) can be used to generate images at the touch of a button that simulate the photographic representation in a deceptively realistic way. This process takes seconds and requires no technical knowledge on the part of the user. Anyone is now able to create realistic images of fictional content—a development that also brings the problem of deepfakes into focus.

This development requires a critical examination. We need to address not only the possibilities but also the risks of this technology and ask ourselves how we want to deal with the increasing manipulability of images.

AG: Let's take a concrete example: you did an exercise with your students at EPFL called "Original & Replica." You set them the task of reproducing a photograph as accurately as possible using prompts in an AI image generator. I find this exercise extremely valuable from an educational point of view because, on the one hand, through this process you understand the mechanisms of these image generators and, on the other, you realize the value of the "original" and the difficulty of reproducing it

PS: In the first module, "My Choice," each student selected four images that either had a special meaning for them or that they felt were outstanding due to their aesthetic quality. The aim was to develop the ability to talk about the qualities of an image and formulate a convincing case for it. Through this endeavor,

the students engaged intensively and analytically with their chosen image. Either personal memories played a role or aesthetic aspects such as composition, coloring, or lighting determined their selection.

In the second module, "AI Reconstruction," the task was to recreate the previously selected original images as accurately as possible using AI image generators. The aim was to use targeted text prompts to create an image that was visually as close as possible to the original.

However, this proved to be challenging. The students had to constantly adapt their prompts in order to achieve a closer approximation to the original and at the same time understand the effects of different formulations (prompt engineering) on the generated image.

It quickly became clear that an exact 1:1 reproduction of the original image using AI image generators with text prompts alone is simply impossible. In addition, there were striking differences in image quality and visual style between the various AI generators. Regardless of the model used, the students also discovered numerous inconsistencies and image errors in their AI-generated images: inconsistent perspectives, faulty depth gradation, problems with fine textures, distorted object shapes, inconsistent light and shadow casts, and faulty reflections on reflective surfaces.

A central aim of the module was therefore to sensitize students to the importance of looking closely at images. They should learn to devote more time and attention to images and consciously pay attention to possible inconsistencies—especially at a time when images are omnipresent, mass-produced, and consumed at high speed on social media.

Fig. 15: / imagine: A tall construction building in a foreboding landscape, in the style of dark, thin steel forms, photo -ar 127:128 (Reference Image: "Brasilia", c. 1958, Marcel Gautherot). Text prompt (above) and AI generated image (Midjourney), 18.10.2023. AR-329: Constructing the View: Built Images, Autumn Semester 2023, EPFL. Student: Darmezin Sidney

Fig. 16: / imagine: A realistic photograph of a thin rectangular black rear-view mirror of an old scooter, a beige sand dune in the background in the distance in front of the observer, the clear blue sky which takes up three quarters of the image, the trail of a plane in the sky, the rearview mirror reflecting another lighter, sunnier beige sand dune, the real dune and the dune reflected in the rearview mirror aligning perfectly as if it were a single dune, no vegetation and no people in the image, the rearview mirror at the bottom right of the image (Reference Image: "Coincidence Project", 2012, Denis Cherim). Text prompt (above) and AI generated image (Adobe Firefly), 20.10.2023. AR-329: Constructing the View: Built Images, Autumn Semester 2023, EPFL. Student: Alix Eggli

Fig. 17: / imagine: Café interior scene viewed from the front. The overall tone is warm. The bottom of the image is in a shadow and the top of the image is illuminated with a warm sunset light. The floor has square white tiles, the wall is composed of a medium dark green panel in its base and mustard yellow paint on the rest of the wall. On the right side of the wall, a landscape painting illustrating a view from a shore with trees on a sea with boats floating is hung. The furniture is 3 sets of rectangle-shaped tables with rounded edges with an off-white color, around each table are 4 dark jean blue plastic chairs with a tall curved back support. Three red bottles of sauces are on top and two spice holders are on each table (Reference Image: "Summerstown", 2019, Niall McDiarmid). Text prompt (above) and AI generated image (Adobe Firefly), 20.10.2023. AR-329: Constructing the View: Built Images, Autumn Semester 2023, EPFL. Student: Sahar El-Zein

AG: This inevitably leads to the question of what an original is. In his famous essay, "The Work of Art in the Age of its Technical Reproducibility," from 1936, Walter Benjamin draws a history of art from the perspective of its reproducibility, from the manual to the technical, lamenting the loss of the aura. Benjamin writes above all against the backdrop of the development of film, to which he attributes a new form of participation, but one that takes place in a state of distraction. This resonates with many phenomena that have intensified through social media. If we now apply this to AI image generation, we are dealing with something new: on the one hand, it is about reproduction—the data that is recorded—and the algorithms that learn from the tasks and produce new images/texts. We are therefore dealing with a peculiar overlapping of production and reproduction, in which the question of an original takes on a whole new meaning, as there are probably many originals that are combined to create something new.

PS: Walter Benjamin was primarily concerned with the fact that the aura of a work of art is inextricably linked to its uniqueness and non-reproducibility—and thus directly integrated into its embedded context. A classic example of this would be a fresco or a wall painting that is tied to a very specific location. However, the situation is different with lithography or photography, which make it possible to produce numerous identical copies or prints of a work of art. These multiplied images can take on very different meanings depending on their context and intended use.

However, AI image generation is difficult to categorize clearly in this context. It is not a pure reproduction tool, as can be seen from the fact that it never generates exactly the same image despite identical prompts. Instead, a new image is created each time, which is difficult to embed in a clearly defined context—as it is not derived from a physical reality, but is based on a complex, data-based construct.

AG: Let's talk about prompts and language. It is exciting that today, images are gaining a new significance, but so, too, are texts. Writing, and especially describing, is taking on a new relevance

PS: In the 1990s, the terms "pictorial turn" and "iconic turn" described the increasing pictorial nature of social communication. To exaggerate and simplify it: the image replaces the word. Today, we are experiencing an interesting kind of reversal of this principle with AI image generation. In order to generate im-

ages, we first have to formulate text input (prompts)—the word once again becomes the central element of image production. This is an exciting development: whereas images used to be seen as a substitute for language, we now use language as a tool to generate images.

Furthermore, images are an extremely democratic medium, as they can be grasped intuitively and understood largely independently of prior linguistic or technical knowledge. AI image generators open up a new dimension of accessibility here: they make it possible to sketch and articulate visual ideas and visions of the future without the need for traditional artistic training.

I see this development as highly positive, as it significantly simplifies the production process and makes it more inclusive. The use of AI image generators requires little prior technical knowledge, which underlines their democratic character. Everyone can develop their own images, concepts, and visions of the future—regardless of their creative skills—and thus actively participate in the creative discourse.

AG: Pictures are an important form of communication, especially when it comes to selling a house.

PS: Yes, it is impossible to imagine architecture today without images. In recent years, however, there has been a strong trend towards increasingly photorealistic project visualization—often at the request of clients or project developers.

However, this type of representation can be problematic, especially in the early design phase. Highly realistic renderings give the impression of an already finished result, leaving little room for open design processes. This can not only restrict creative development, but also mean that alternative approaches are no longer even considered.

I therefore encourage my students to try out different aesthetic approaches in order to expand their digital visual vocabulary and develop a more conscious approach to images. Instead of being guided by a single, supposedly "correct" way of representation, they should learn to use images as tools for thinking and designing—and not just as a final form of presentation.

AG: In this context, there is a lot of talk about creativity, and we also have two texts on this topic in the book. Creativity is seen as the only human ability that cannot be taken over by AI, because the technology can only reproduce.

PS: From my point of view, it is quite clear: AI does not replace creativity but remains dependent on humans as authors. It needs an idea, an intention, a creative decision, which is then further developed with the help of an image generator. Without conscious control, a concept, or an artistic question, the generated image remains just a random arrangement of pixels.

AG: One topic that comes up again and again in this context is the speed of processes. We are dealing with an acceleration that is overwhelming many people. How do you see this development? I imagine that your work on the picture is a very lengthy process; is this now accelerated with AI?

PS: Yes, many things have indeed become faster—that's true. But this acceleration also has positive sides. Today I have a much greater variety of image material and variations at my disposal than was previously available. This opens new possibilities and considerably expands the creative scope.

However, this unlimited variety can also seem overwhelming. I would see it less as overwhelming and more as a constant surprise at how many possibilities arise in a very short space of time. The challenge lies in making the right decisions and not getting stuck in the flood of options.

Ultimately, for me, it's about using the technology sensibly—as a tool that supports my creative process, but doesn't determine it. I use AI where it helps me to explore things faster without losing focus on the content.

AG: So you have adapted a new tool to your old working method.

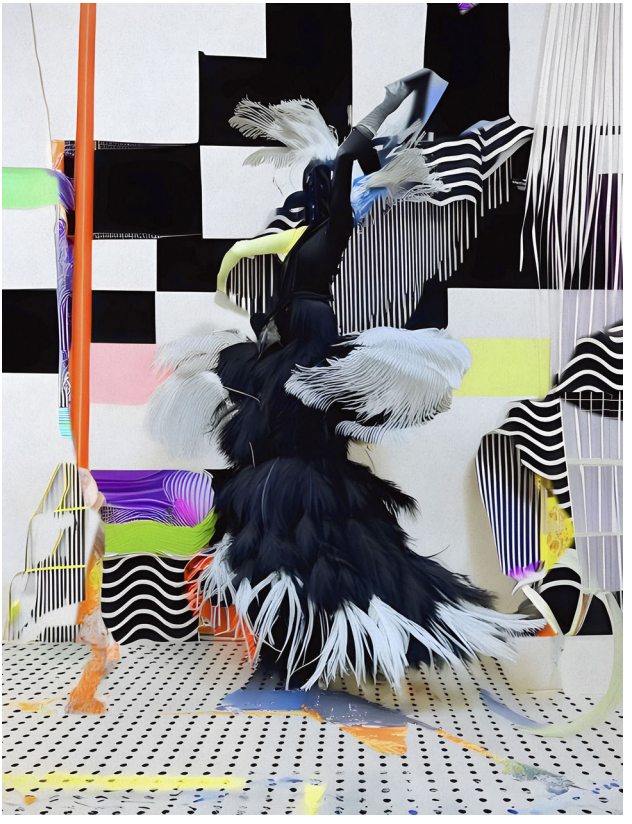
PS: Yes, I think we can say so. In the past, my work may have taken more time in terms of craftsmanship than it does today. In this sense, the work actually loses some of its artisanal character, as I can now generate fragments in a more targeted and efficient way. The creative process shifts more to the curatorial and conceptual level, which is about making the right decisions from the multitude of possibilities.

AG: But that also means, when we come back to teaching, that students have to understand this decision-making process.

PS: Yes, exactly. Students must learn to consciously cultivate the decision-making process and develop cultural depth—a skill that is always linked to their own biography, perception and experience. It is also about gaining confidence

in this role. The ability to make a choice and to make well-founded decisions is not self-evident—and is otherwise rarely taught systematically. It is therefore crucial that they learn to reflect on the medium and the image in order to sharpen their own visual language and deal more consciously with the flood of possibilities.

Fig. 18: Philipp Schaerer, from the series "Crossbreeds – Imaginary Still Lifes", 2024



AG: Where do you see the potential of AI in architecture?

PS: AI image generation is ideal for initial image approximations, concept images, tests, and variant studies. But architecture is precise: tied to a specific location, and usually thought through down to the smallest detail. It is simply impossible to precisely describe and fully control all the image-specific elements of architecture with text prompts alone.

It is true that it is now possible to upload images in addition to text instructions and link these to the prompts, which improves the result to a certain extent. However, even this method cannot meet the requirements of an architect's planning precision.

The further development of AI models in a direction in which 3D data of the architecture and the context with material assignments are uploaded instead of text prompts to describe the architecture seems more promising to me. Text-based input would then only be used to define stylistic image features. I am convinced that this is only a matter of time—initial approaches to such models already exist, even if they are not yet fully developed.

I have always been interested in hybrid image constructions—in other words, images with breaks and varying degrees of abstraction. These include, for example, photographic fragments that at first glance appear to be "illogically" integrated into an ensemble or interwoven with more abstract pictorial elements.

In general, I believe that we should take a step back in architectural representation—especially when it comes to communicating ideas and designs. We need to move away from the exclusively photographic visual language that increasingly serves as the standard for communicating concepts and possibilities.

This is precisely where I see great potential in AI image-generation programs: they could help us to explore and test new image approaches and more abstract aesthetics.

AG: Last but not least, I would like to address the question of authorship. What about the authorship of an AI-generated image when, in principle, potentially thousands of other people's images are recorded there in the form of data?

PS: The issue of authorship and copyright is important and real. A current example is the lawsuit filed by the *New York Times* against Microsoft and OpenAI because they believe that their copyrighted content was used on a large

scale without permission. However, only large media companies can afford such lawsuits—many artists and creatives do not have the resources to legally defend themselves in this way.

It is therefore all the more important that we talk to students about the ethical aspects of AI in image production. Transparency plays a crucial role here—both in terms of the tools used and the creation processes. Today, it is more important than ever to declare the techniques and methods used to create images. This not only reflects the respective authorship but also reveals how and by what means an image was created. At a time when digital image production and AI generation continue to advance, it is becoming increasingly difficult to distinguish between handmade, photographed, or synthetically generated images. The open labeling of the tools used creates transparency and enables a critical examination of the image source and production.