

ELMAN MANSIMOV



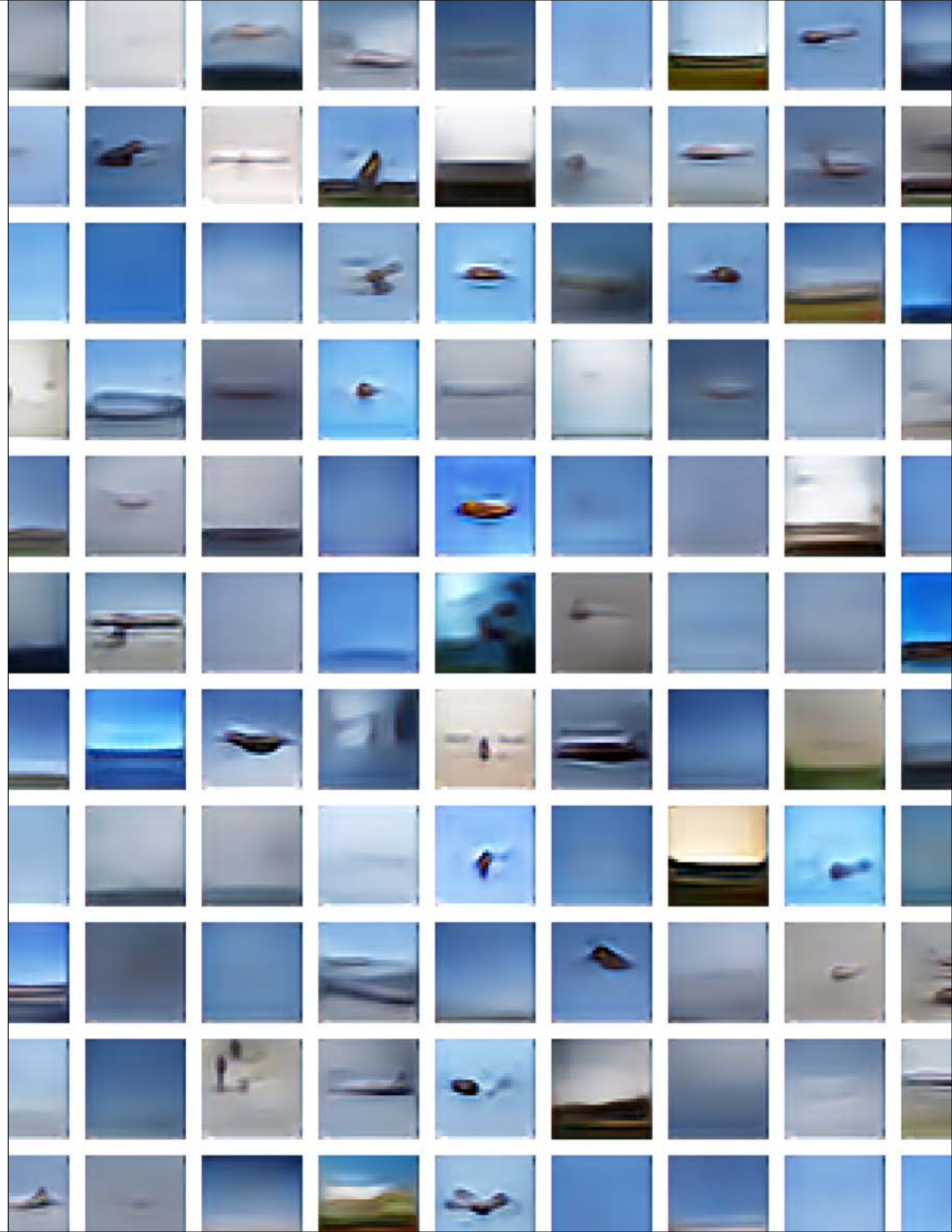
Self Portrait by Elman Mansimov

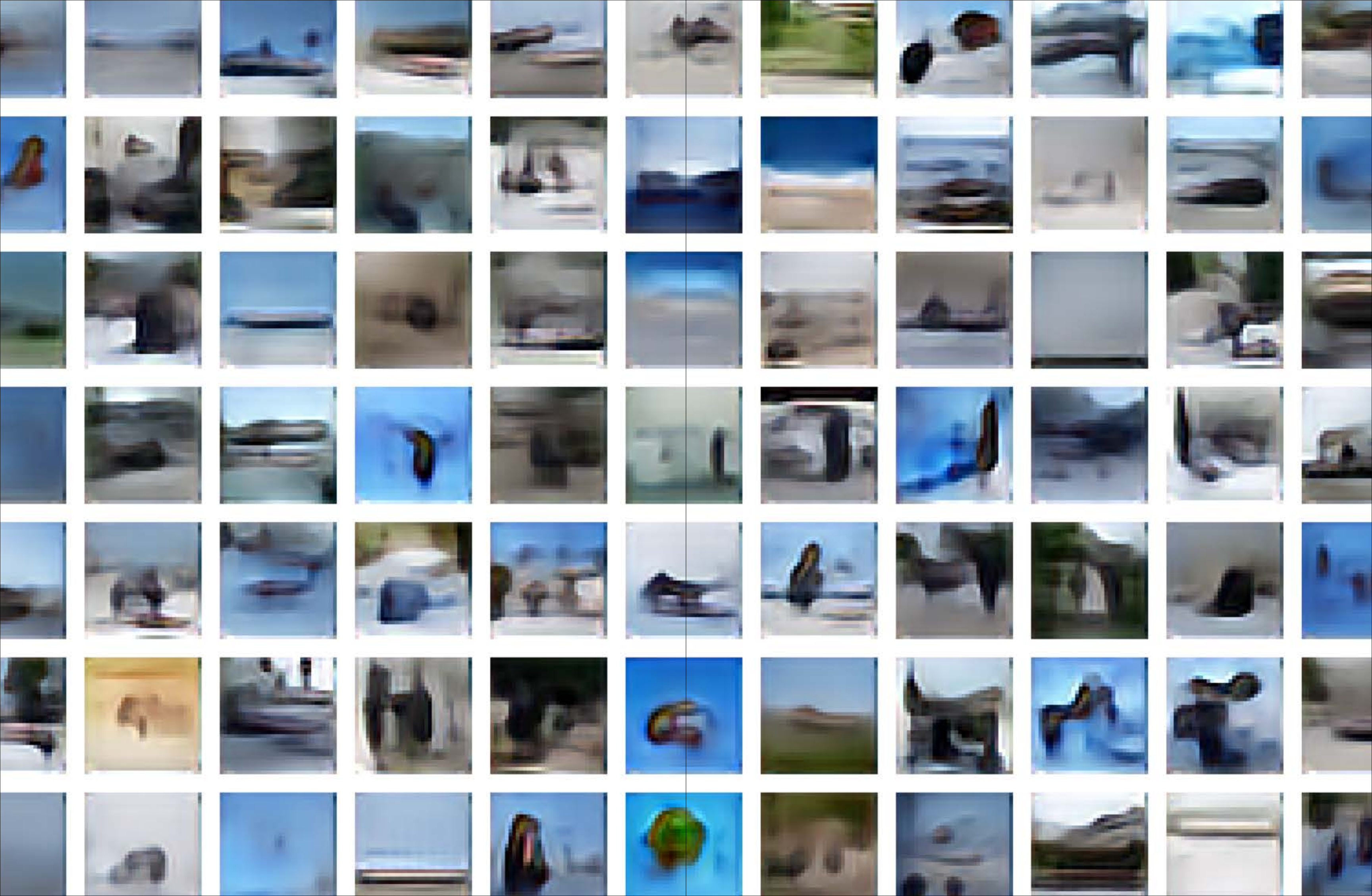
THE SCIENTIFIC METHOD

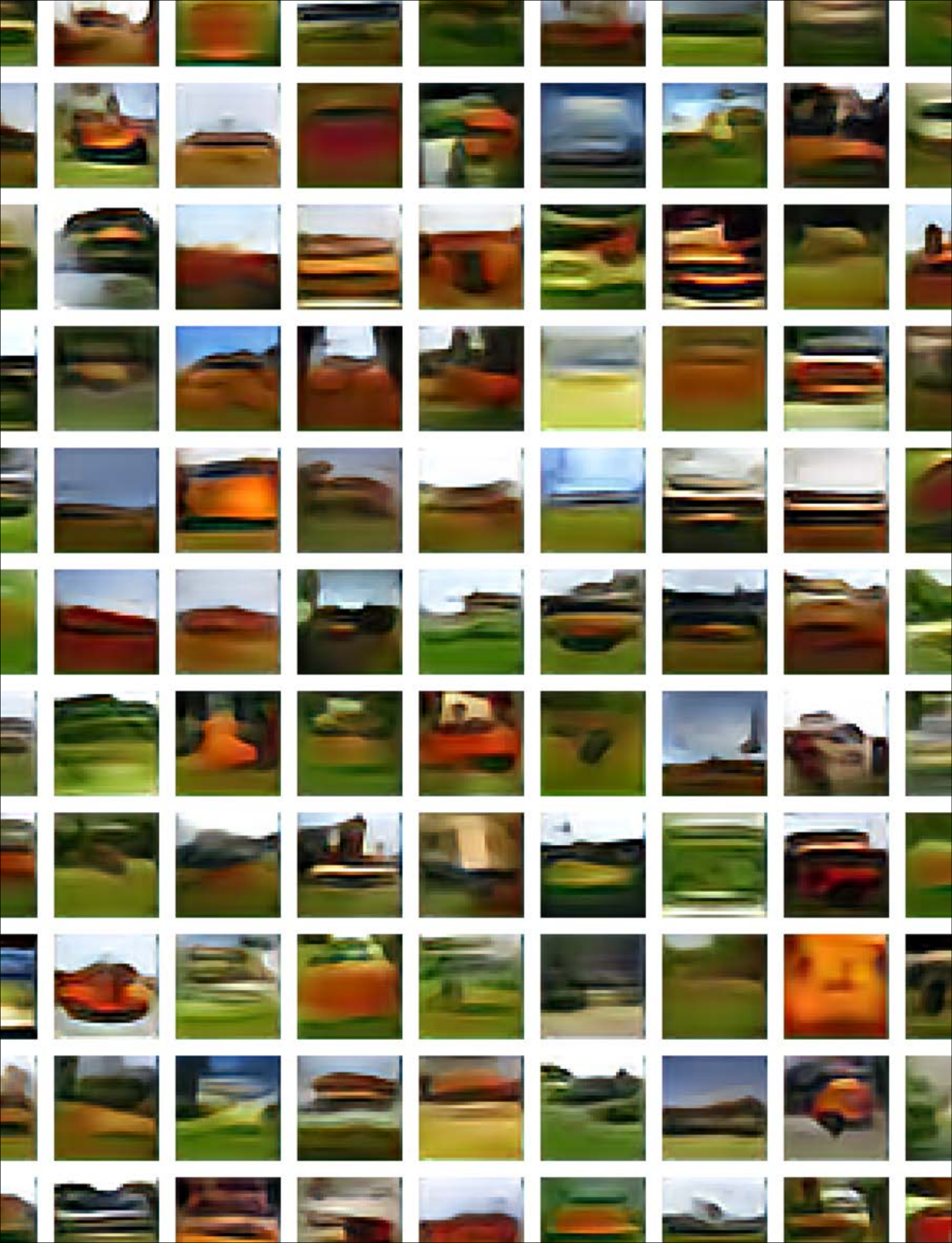
STEVE MILLER: Your presentation at Fellowship Gallery at Paris Photo revealed a history unknown to me. That being, revealing the origin of text prompt to image in 2015. What made you think that words could make an image? **ELMAN MANSIMOV:** I was always looking at it from the computer science perspective. At the time, I was a computer science student at University of Toronto. I started getting involved with a machine learning group that was training neural networks to take an image and classify it or take an image and write a caption about it. And in principle, there was nothing preventing us, conceptually, to take the same framework that does image captioning or image classification and just reverse it. This was a research project. It was never meant to be an art. I felt like it's definitely going to be interesting for the machine learning research community and let's just work on it and see where it goes. **STEVE:** So it was just

pure curiosity that you could do the reverse of scanning images; to actually give a set of instructions that would create an image? **ELMAN:** Yeah, it's a mix of curiosity and also a mix of experience because I started working in the machine learning lab. My first project was to predict the next frame in a video. So if you've seen all these language models like a GPT from OpenAI, the way they work is they try to predict the future, they try to predict the next word in the sentence, and essentially by doing that you learn the knowledge of the world. Like you know that, let's say the word, peanut is likely followed by peanut butter. We could successfully try to predict, by that time, to predict how the digits move around the screen. I

Elman Mansimov, Opposite: (Detail) *Process*,
An airplane flying off into the distance on a clear day, 2015;
Following spread: (Detail) *Process*, A herd of elephants
flying in the blue skies, 2015.







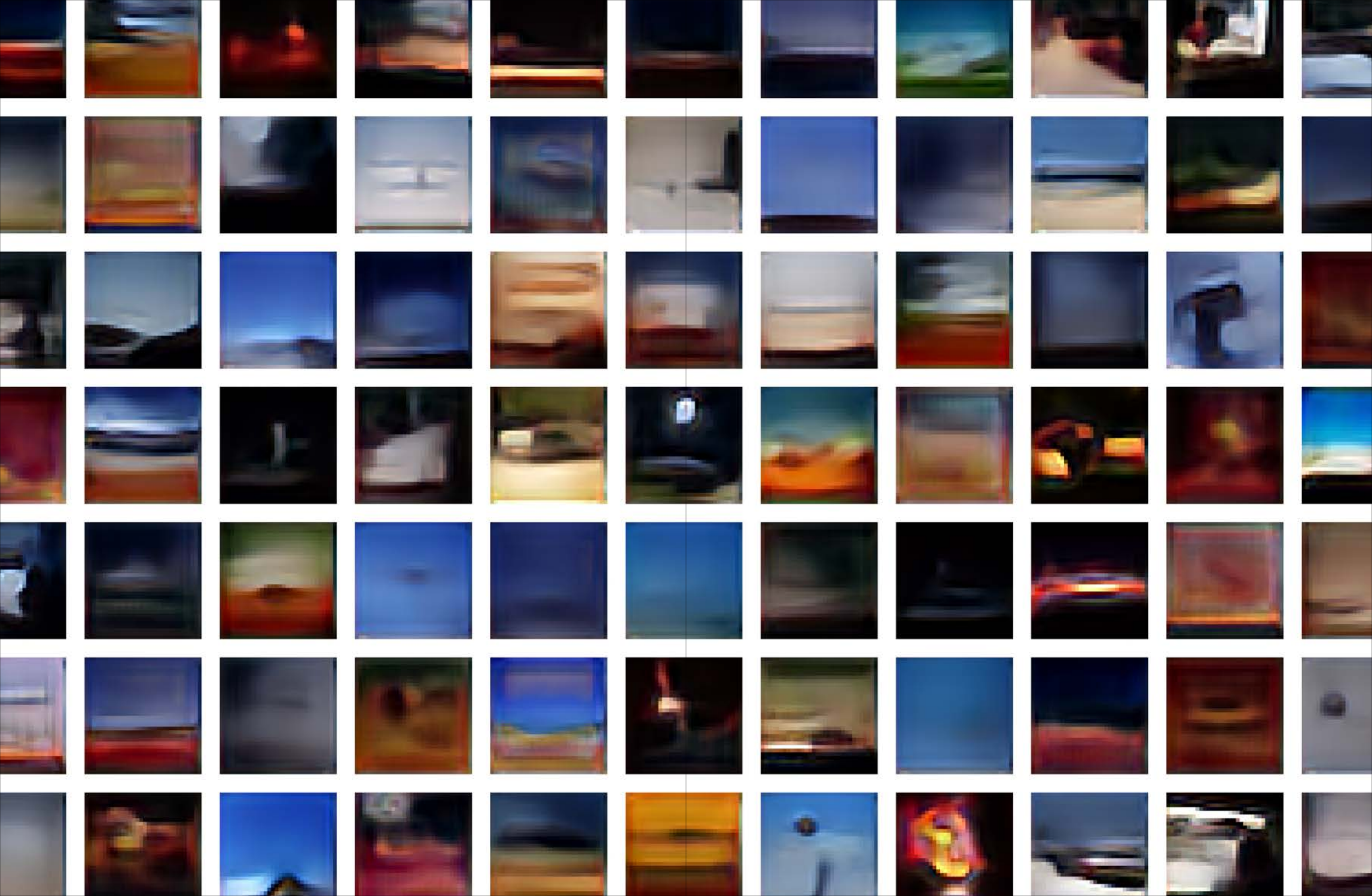
had some success doing it and let's just try reversing the image captioning process. And that's how it was. **STEVE:** *Well, that sort of leads to the next question. The results are fascinatingly primitive by today's 2023 standards of DALL-E and let's say Midjourney for example. You're looking at the future, which is out of focus in 2015. What was your response to seeing these initial out-of-focus images?* **ELMAN:** That was amazing. I remember it was either the end of 2021 or early 2022 when DALL-E 2 came out. I was pleasantly surprised and I was very happy to see it. To see how far it went from 2015 and especially putting it in the perspective of how things are now. It's not just the accomplishment of people behind Midjourney, DALL-E or me back in 2015 but, I feel like it's a collective work of the world in a sense because hardware got better, the GPUs got better, and the algorithms to train these neural networks got better too. People figured out how to train these neural networks better. The software improved. The thing is how technology advanced and how it's kind of helping each other to get to the point that we are now. **STEVE:** *Did you even imagine the sophistication of today's renderings?* **ELMAN:** Oh yeah. I think that actually, my intention was to be as sophisticated as we are now in terms of images. But I was also disappointed at the same time because I thought that maybe, for my first try, I would be able to create something as advanced and as realistic as today. Of course, I failed in that respect, but I also succeeded in other aspects. **STEVE:** *So photography has traditionally been defined as light passing through a lens and a new definition has to include this text prompt model. And how do you define photography and its evolution?* **ELMAN:** It's a very good question because, as you said, photography is, in the traditional aspect, you have a camera and then you have a light passing through the lens. Then maybe you post-process it in some way, you work on it and then you have a picture. It's also a phenomenon that happens in the physical world. You come in with a camera and then take a picture of the real thing that's happening in the real world versus with a text to image, with a generative, generating images so, it is a "kind" of photography. I can write a prompt now with a Midjourney and say, "Generate me this lens, Sony camera, that kind of lens of, let's say, a building or let's say a person posing in front of the camera." But I think we need a new definition of that because I wouldn't necessarily call it photography. Either we need to expand the definition of photography or we need to call it virtual photography or AI photography because it's not in the sense the physical photography that we do now, right. It's AI generated pho-

Elman Mansimov, Opposite: (Detail) *Process*,
A yellow school bus is walking across a green grass field, 2015;
Following spread: An airplane flying off into the distance at night, 2015.

tography. **STEVE:** *You followed this trail of text prompts with the idea that you were a scientist, but now you're showing at Paris Photo and you're being contextualized as an artist. Do you associate your work with science and computers with the perception of working as an artist?* **ELMAN:** You know, I associate this work as the work of science and research that by accident became an art. The original intention was never to be an art. I never even thought that in 2015 I would be in the Paris Photo. So it's a very pleasant accident. The images are now kind of a

I ASSOCIATE THIS WORK AS THE WORK OF SCIENCE AND RESEARCH THAT BY ACCIDENT BECAME AN ART

piece of history, a piece in the museum that documents the thing that people used to have. The field has exploded. And by being on the right train in some way, that's going in the direction that is exponentially up. It ended up being a work of art, so now I consider it to be both. I consider it to be kind of a research work from its origins, but now kind of a work of art and a piece of history at the moment. And then for probably indefinite future. **STEVE:** *Is art science a false dichotomy or how would you merge those two? Or what would be the new definition if there is one?* **ELMAN:** I think I would say creation or creators because I feel like at the end of the day, scientists and artists are creators. Even in the case of machine learning, deep learning, a lot of the things are not necessarily scientific, as in, we have them in a proof. It's like a natural thing in the world that we study, a natural phenomenon in a lot of the things that we create. And I think a similar thing with the art we create. We create photos, paintings, and drawings. The right way to call both artists and scientists is creators. Certain people who create things become lucky because they stand the test of time or end up being a very important thing in the world and in history. So yeah, I would call those creators. **STEVE:** *With AI, instead of looking at the world through a lens, you can imagine and replace the physical eye with the mind's eye, or just plain thinking. It's kind of revolutionary. There was this intermediary thing that was a camera. Now the intermediary thing is a computer, but the fluidity of the computer far out-*



weighs the camera. Any thoughts about that? **ELMAN:** The way I think about it is, I feel like the human mind is the first thing. If you think from the biological device, you think from that perspective. We have a thought in our heads and might try to communicate it through language. Or you go to the cave, you see how cavemen tried to communicate their thing with art. They would draw the hunting process and their life there and the first pieces of art. If you think about it, over the course of history, we tried to scale it from books to printing media to paintings and drawings to computers and manufacturing. Essentially, in every de-

IN EVERY DECADE, EVERY CENTURY, WE CREATE A DEVICE THAT HELPS SCALE OURSELVES, SCALE OUR IMAGINATION, SCALE OUR THOUGHTS

cade, every century, we create a device that helps scale ourselves, scale our imaginations, scale our thoughts and then distribute it in a much, much wider audience, in a much wider way. So that's how I personally think about photography, computers, or any kind of device. It's just a natural evolution of creation and then creative thought. The human mind is supposed to create things and innovate. The camera is just one way of doing it. **STEVE:** *Al relies on the camera because all those photo images can be accessed in a database. It's valid to call it a form of photography because it's collaging all the photos that have ever been made. The camera's still there, it's just being reorganized. The computer is not replacing the camera because it's scraping camera images. So it's like this generative extension of the camera.* **ELMAN:** Yep, that's actually a very good way of thinking about it. And then you're not only extending the cameras, the photos of the camera, you're extending so many of the pictures in the world from all kinds of domains, all kinds of mediums. And then you have a human touch. People putting in their biases on how they build these models to generate those images, I think that's also a part of it. **STEVE:** *So just as photography marginalized the need for representation that stimulated the possibility for abstraction, AI transformed the same traditional role to take a photograph, which anyone can do, but now anyone can also imagine a photograph. So have you thought about the glut of the proliferation of AI images and the next impossibility of sorting through them? For example, in your early experiments, like your presentation at Paris Photo, what was*

your criteria for editing, or we would say, curating the images into smaller groups? **ELMAN:** That's a good question because in principle, even back then, I could have just generated millions of images. Nothing was preventing me. Back then, the intended audience was scientists. My criteria was what are the interesting prompts and what are the resulting images that I can select to demonstrate the promise of the model. I guess it's similar to an artist or photographer. When you're creating something, you are creating something for others to feel something, to be inspired, to be fascinated. That was the criteria. The dataset back then was, let's say, 82,000 images from the Microsoft COCO, and they all were real images of the real world. I needed to convince people that the model is different from the Google search. It's taking unrelated concepts and then combining them together in a way that looks realistic but you wouldn't find on Google search. Why don't I just combine different concepts from the data-

set together into a prompt and then generate it? AI can do what humans with cameras can't, because cameras are always restricted to the real world, to the physics of the real world. **STEVE:** *But are those larger grids in Paris Photo? Had those already been predetermined back in 2015?* **ELMAN:** Yeah, that was all 2015 images. **STEVE:** *So have you shown your work in a fine art context previous to Paris Photo?* **ELMAN:** No, never. That was actually the first time that it was shown in a fine art context. **STEVE:** *Did you do anything new for Paris Photo in terms of aesthetic curation?* **ELMAN:** That's a very good, interesting question because I never think about curation because I'm not an artist. And of course, I don't know anything about exhibitions, about the art world. I'm just a casual observer of it. I go to the museums in New York, like the Guggenheim, and then paint and draw in my free time. Experts pitched to me, as there's a historical significance of this work, especially in this time when generative AI and photography AI is taught. So, it made sense to exhibit this work because it's now a piece of history. Paris Photo gave a bigger, broader context of this work to artists and photographers who attend Paris Photo. Now it's no longer just a piece of history or a piece of science, it's a piece of art as well. **STEVE:** *I think I'm catching you talking like an artist.* **ELMAN:** As a creator, yes. As a creator.

Elman Mansimov, (Detail) *Process*,
A yellow school bus is walking across a green grass field, 2015.

