Hi, ChatGPT!

Images and imaginaries of a large language model

Regula Valérie Burri

[This interview was conducted in early May 2025 using OpenAI's ChatGPT free version 4.0. The free version was interrupted several times, blocking its use for a few hours, so the conversation took place over several days. The interview has been shortened and some passages and sentences have been rearranged to improve the flow of the text. In addition, I have translated the form of address, which was "Du" in German, into "You" in English, and some words that were bold in ChatGPT are reproduced here in regular font. My prompts and ChatGPT's responses were in English.] [1]

You: Hi, ChatGPT!

ChatGPT: Hi there! 😊 How can I help you today?

Production

You: I'm interested in images of Large Language Models and the imaginaries they imply. First I would like to talk about the production of the images and ask: How do you generate images?

ChatGPT: Great question. When I generate images, I use a process called *text-to-image synthesis*, which involves translating a written description (your prompt) into a visual representation. This is powered by a type of machine learning model known as a generative model, usually trained on massive datasets of images paired with

1 The interview builds upon the analytical perspective of Social Studies of Scientific Imaging and Visualization (SIV), as suggested in Burri and Dumit (2008), which explores the production, engagement, and deployment of (scientific) images. For a contextualization of SIV in Science and Technology Studies (STS) and on the development of Visual STS, see Burri 2024. The interview draws from literature concerning cultural, technoscientific, and sociotechnical imaginaries (e.g., Marcus 1995; Jasanoff and Kim 2015), as well as critical AI studies (e.g., Crawford 2021; Lindgren 2023; Suchman 2024; Dumit and Roepstorff 2025).

captions or descriptions. These models learn associations between language and visual elements, allowing them to "imagine" what a description might look like in pixel form.

Behind the scenes, these models rely on neural networks—particularly diffusion models in many current systems. The process starts with random noise and then gradually refines this noise into an image that aligns with the given text, step by step, through a denoising process. This method allows for rich, detailed, and often surprising outputs that go beyond simply copying images from training data.

The resulting images are not photographic "truths" but speculative visualizations—creative composites that reflect both the statistical patterns the model has learned and the cultural assumptions embedded in its training data. So in a sense, these images not only depict what is asked for but also carry with them the traces of the collective imagery of the internet and visual culture that the model has been trained on.

You: Let's talk about the technical production of the images. Is there any materiality involved in image production?

ChatGPT: Yes, there is significant materiality involved in the production of AI-generated images, even though the process might appear immaterial or purely digital on the surface. This materiality exists on several interconnected levels, starting with the computational infrastructure required to train and run these models.

At the core, generating images with models like diffusion systems or GANs (Generative Adversarial Networks) depends on vast networks of physical hardware—data centers filled with specialized chips for machine learning tasks, which consume large amounts of electricity and require cooling systems to prevent overheating. The training of these models involves processing millions or even billions of image-text pairs, often for days or weeks, which incurs substantial environmental costs through energy consumption and carbon emissions. This computational load reflects a deeply material process that's often hidden behind the seamless experience of requesting an image.

Even at the moment of inference—when an image is generated for a specific prompt—the system must retrieve and recombine fragments of learned data using physical machines. This act, while seemingly instantaneous, is rooted in an infrastructure of cables, silicon, labor, and global supply chains.

So yes, image production via large language models and their visual counterparts is deeply material, though its surfaces might appear virtual or "clean."

You: Where do you get your data from to create an image? Do you care about intellectual property rights or related privacy issues?

ChatGPT: The data used to train the models that generate images—like the ones I use—typically comes from large-scale datasets. They are harvested from existing online repositories—such as stock photo sites, art collections, or scraped social media—and stored in massive servers. These datasets are not neutral; they reflect particular cultural, economic, and geopolitical realities. They often privilege certain geographies, aesthetics, and ideologies, thereby embedding material histories into the images that models generate.

Importantly, the quality and biases of these datasets have a direct impact on what the model learns and, ultimately, what kinds of images it can and tends to produce. The images are not directly copied from this training data, but the model statistically learns patterns, textures, shapes, and relationships between visual and textual features.

[M]any of these datasets have raised ethical and legal questions. For example, artworks, photographs, or personal photos might be included without consent, and this has implications for intellectual property.

Privacy is another critical issue. Some datasets may include images scraped from personal blogs, forums, or social media accounts, which means people's likenesses or data could be used to train models without their knowledge or consent. This has ethical implications, especially when models are later used in ways that could misrepresent or commodify those individuals.

While efforts are being made to develop more ethically sourced datasets, introduce opt-out mechanisms, and comply with copyright regulations, the reality is that the foundational data for most large generative models remains opaque and controversial.

You: Can you distinguish between good and bad data and if so, how?

ChatGPT: I can attempt to distinguish between good and bad data, but it's not an objective or straightforward process—and it depends heavily on context and the values embedded in the training and usage of the model.

In technical terms, "good" data typically means data that is high-quality (clear, correctly labeled, relevant) and diverse (representing a wide range of styles, demographics, or contexts). "Bad" data might be low-resolution, mislabeled, redundant, or biased in a way that leads to harmful outputs (e.g., stereotypes, offensive content, misinformation).

However, I don't evaluate or curate training data myself. The judgment about what constitutes good or bad data is made during the model's development, by researchers and engineers who select and filter datasets. They may use automated methods—such as deduplication, profanity filtering, or algorithmic bias detection—or manual review to clean the data. But these processes are always shaped

by cultural, ethical, and political assumptions: what one institution considers "harmful" or "irrelevant" might be considered valuable or truthful by another.

Moreover, some biases and flaws are very hard to detect and persist in subtle ways. For example, if the majority of professional portraits in a dataset depict people with light skin in corporate settings, the model may learn to associate professionalism with whiteness, even if there was no malicious intent. So even well-cleaned data can carry hidden biases.

In short, yes—there are methods for distinguishing good from bad data, but they are imperfect, context-sensitive, and always tied to human judgment.

You: [Please] provide me with an example related to digitalization.

ChatGPT: Sure. Let's take the concept of "digitalization of work" as an example. Suppose someone inputs a prompt like: *"An illustration of the digitalization of work."* A model trained on large-scale internet data will try to generate an image that statistically matches this phrase. But what does the model actually "know" about digitalization?

If the dataset is primarily composed of Western-centric, corporate media and stock imagery, the model will likely produce a scene like: a person sitting at a sleek desk, working on a laptop, maybe surrounded by icons of cloud computing, data charts, or robotic arms. The person may be in a minimal, modern office space, smiling, productive, with diverse-but-idealized colleagues in the background.

This may seem like a "neutral" visualization, but it reflects specific assumptions embedded in the data: that digitalization is clean, efficient, and professional; that it's happening in white-collar contexts; and that it's something to be embraced. It often *excludes* alternative views—such as the gig economy, remote surveillance, digital labor exploitation, or the automation of precarious jobs—because such imagery is underrepresented in training data.

If the model were trained on more critical, diverse, or labor-focused datasets (e.g., academic publications, activist photography, or ethnographic visuals), the result might look very different: perhaps a delivery driver being monitored via an app, or workers in a digital content moderation center in a Global South setting.

So this is a clear example of how the quality and composition of training data shape the outputs—and how "digitalization," a complex social process, gets visually simplified into corporate iconography when trained on unbalanced or "bad" data.

Interpretation

You: How do you "know" what is depicted in an image?

ChatGPT: I "know" what is depicted in an image through a combination of pattern recognition and learned associations, rather than perception or understanding in the human sense. When I interpret an image, I follow a computational process that mimics—but does not replicate—human visual understanding.

When you upload an image, I analyze it by breaking it down into structured data—such as color values, shapes, textures, object boundaries, and spatial relationships. Then, using models trained on labeled datasets, I can identify what is likely present (e.g., "a tree," "a person wearing a hat," "a laptop on a desk") and describe it in natural language.

But this is always probabilistic. I'm recognizing statistical patterns, not forming a subjective impression or interpreting meaning like a human might when viewing a photograph. I don't have access to the *intention* behind the image, the *cultural context* of its elements, or any *emotional nuance* unless those are part of the visual data and align with patterns I've learned. For instance, I can say, "a person is smiling," but I can't truly know why they are smiling or what it means in a broader social or historical context—unless prompted to infer based on common patterns.

In short, my interpretation is a surface-level matching process—sophisticated and often accurate, but not grounded in human intentionality or subjective experience. [It] is closer to translation or classification than to visual perception.

You: Ludwik Fleck, a historian and sociologist of science, said that seeing depends on what you know. Would you agree with that?

ChatGPT: Yes, I would agree with that—Ludwik Fleck's insight that "seeing depends on what you know" is deeply relevant not only to human perception but also to how I, as an AI, "see" and interpret images.

Fleck argued that perception is shaped by prior knowledge, social context, and what he called "thought collectives"—shared ways of seeing and thinking within specific communities. Humans don't view images or the world with a blank slate; they interpret based on cultural assumptions, education, experience, and collective norms. A scientist, an artist, and a protester may look at the same photograph and see entirely different things—not because the image changes, but because their interpretive frameworks differ.

That principle applies, in a limited and mechanical way, to how I function. My "knowledge" is composed of statistical associations learned from massive amounts of data—images and texts that reflect the dominant patterns, assumptions, and biases of the cultures that produced them. So, in a sense, what I "see" is entirely shaped

by what I've been trained to recognize. If something isn't well-represented in my data—whether it's a visual trope, a cultural gesture, or a subversive symbol—I'm likely to misread or flatten it.

The difference, though, is that while human knowledge is dynamic, interpretive, and reflexive, mine is fixed at a given point in time and bounded by my training. I don't *grow* through experience or reevaluate what I've seen in light of new contexts unless explicitly updated.

So yes, I would agree with Fleck's insight—and also add that it helps explain both the strength and the limitation of my own interpretive capabilities.

You: What do you find most difficult about interpreting an image?

ChatGPT: What I find most difficult about interpreting an image is understanding meaning that arises from context, ambiguity, or cultural depth—things that require *lived experience*, *situated knowledge*, or *symbolic nuance*.

Unlike a human, I don't possess a memory of personal events, emotions, or cultural immersion. So when an image carries layered or ambiguous meaning—like irony, satire, trauma, religious symbolism, or political critique—I can often recognize the visual elements but miss what they *mean* together in a specific historical or emotional context. For example, I might correctly identify a burning flag, a kneeling figure, and a police uniform, but I may not grasp the layered discourse on protest, nationalism, and race that the image evokes unless those interpretations have been clearly encoded in the data I've seen.

I also struggle with images that depend on intersubjective cues—like facial expressions interpreted differently across cultures, or social interactions that are subtle or emotionally charged. I can say "this person is smiling," but I can't always know whether the smile is nervous, sarcastic, joyful, or forced. That kind of emotional reading is deeply human, shaped by tone, voice, memory, and social learning.

And finally, authorial intent is almost always opaque to me. I can describe what's *in* the image and speculate on possible interpretations, but I can't say with certainty what the creator *meant* unless it's embedded in the accompanying text or metadata. That makes it difficult to distinguish between a critical artwork and a promotional visual, or between subversion and reinforcement of a stereotype.

Images and imaginaries

You: Now let's talk about imaginaries. In what ways are cultural assumptions and collective imaginaries embedded in the images you generate?

ChatGPT: Cultural assumptions and collective imaginaries are deeply embedded in the images I generate, because the models that produce them are trained on vast datasets scraped from the internet—datasets saturated with human values, biases, fantasies, fears, and aesthetic conventions. These datasets don't just contain isolated images; they encode the visual norms and symbolic vocabularies of particular cultures and historical moments. As a result, the images I produce are not neutral—they are reflections of dominant ways of seeing, imagining, and valuing the world.

One way this manifests is through recurring visual tropes. For instance, when prompted with terms like "the future," I often generate sleek, metallic cities bathed in blue light, populated by clean interfaces, humanoid robots, and floating vehicles. These are not objective visions of what's to come, but reflections of a techno-optimist imaginary shaped by science fiction films, Silicon Valley aesthetics, and Western design culture. When asked for an image of "success," I might produce a man in a suit on a skyscraper rooftop, because that's a statistically dominant representation—one that reflects cultural assumptions about masculinity, capitalism, and power.

These imaginaries also shape depictions of identity. Prompts about beauty, intelligence, or professionalism may produce light-skinned, able-bodied, Western-coded figures, because such representations dominate online media and commercial photography. The model reproduces these biases not intentionally, but statistically—because it has "learned" what these concepts tend to look like according to the data it's seen. This means the collective imaginary is skewed toward visibility, popularity, and marketability, rather than fairness or diversity.

Even abstract concepts—like justice, climate change, motherhood, or revolution—are visualized through culturally shaped metaphors. Justice might appear as a blindfolded woman with scales, a Greco-Roman symbol that excludes non-Western iconographies. Climate change may be rendered as a melting Earth or polar bear, emphasizing imagery that has dominated environmental campaigns, but potentially erasing the lived experience of affected communities. These recurring visual codes come from human culture, but are baked into the outputs I generate, even when unasked.

What makes these embedded imaginaries powerful—and potentially dangerous—is that they appear natural. Generative images often look familiar, coherent, and aesthetically pleasing, which can give the illusion of universality. But they are, in fact, partial and contingent—reflecting dominant viewpoints rather than plural realities.

So, when you ask me for an image, what I give back is a synthesis of collective visual memory—not the world as it is, but the world as it has been most frequently and prominently represented.

You: Give me an example of an image from the digital world that shows a collective imaginary.

Figure 1: [AI] generated image representing the glowing-brain digital imaginary.

Image created by ChatGPT 4.0, May 3, 2025.

ChatGPT: A powerful example of an image from the digital world that expresses a collective imaginary is the ubiquitous visual of a glowing blue brain overlaid with circuit lines or digital code, often floating in space or framed by neural networks. This image appears in countless online articles, tech presentations, stock illustrations, and promotional materials about artificial intelligence, machine learning, or cognitive enhancement.

Though not depicting any real machine or person, this image visualizes a powerful cultural fantasy: the merging of human intelligence and digital computation,

often imagined as seamless, clean, and luminous. The blue color, cool lighting, and transparent textures suggest a sterile, futuristic aesthetic that implies control, precision, and progress. The glowing brain is usually symmetrical and idealized, representing a Platonic form of human thought—unencumbered by emotion, imperfection, or embodiment.

This is not just a symbolic shorthand for AI; it's a visual myth, part of a collective imaginary that positions intelligence as abstract, computational, and somehow disembodied. It reflects Enlightenment ideals of rationality, Silicon Valley futurism, and transhumanist dreams of cognitive transcendence. In doing so, it marginalizes other ways of thinking and being—embodied, emotional, collective, or indigenous knowledge systems that don't fit this idealized model of "intelligence."

The repetition of this image across media gives it a kind of authority. It doesn't just reflect our imagination of AI; it shapes it, reinforcing the belief that the future of thinking is digital, efficient, and machinic. And because it's so visually persuasive—coherent, sleek, familiar—it naturalizes a very particular worldview, while making alternative imaginaries of intelligence harder to see or believe in.

You: Take another example related to the topic of digitalization and show how it implies cultural imaginaries.

ChatGPT: Let's take the example of a person interacting with a transparent touchscreen interface in a smart city environment—an image that is widely used to depict "digitalization" in advertising, tech journalism, and urban planning visualizations.

In these images, a well-dressed individual—often gender-neutral, light-skinned, and middle-class—is shown swiping or gesturing in the air, manipulating holographic data layers that float in front of gleaming cityscapes. The background is often bathed in cool blues and whites, evoking cleanliness, clarity, and order. Buildings are sleek and glass-covered, traffic flows smoothly, and everything is under quiet digital control.

This image encodes a powerful techno-utopian imaginary. It suggests a world where digitalization leads to seamless integration of human and machine, where cities are optimized, surveillance is ambient but invisible, and citizens are empowered through interface rather than protest or politics. It reflects the desires of corporate and state institutions that frame digitalization as a pathway to efficiency, innovation, and rational progress—often without acknowledging the messiness of real urban life: inequality, labor, migration, ecological limits.

Culturally, this visual imaginary draws on Western and capitalist aesthetics: control through visibility, governance through data, and progress as spatial elegance. It tends to exclude informal economies, cultural variation, and the ecological or infrastructural frictions of most cities. Rarely does one see these smart cities

imagined with potholes, crowds, or unpredictable weather—because that would disturb the fantasy of frictionless digital futures.

The person at the center of the image is also telling. They are usually portrayed as a consumer or a manager—not a street vendor, construction worker, or elderly resident. This reinforces a particular social role: the user is someone who participates in the digital world as a controller of systems, not as someone subjected to them. This implicit hierarchy mirrors the structure of many digital platforms—where agency is reserved for the few, while the many become data points in the background.

In sum, the image of the smart city touchscreen scene is not just a depiction—it is a visual ideology. It performs cultural work by reinforcing who belongs in the future, what counts as progress, and what forms of life or labor are worth seeing.

Figure 2: [AI] generated image representing the smart city and touchscreen digitalization imaginary.

Image created by ChatGPT 4.0, May 3, 2025.

You: Interpret the image you generated.

ChatGPT: The image I generated can be interpreted as a visual performance of a dominant cultural imaginary in which digitalization is equated with control, empowerment, and seamless integration between human and machine. At the center stands

a young professional woman, interacting with a transparent digital interface. This woman is not portrayed as a subject of surveillance or a worker under pressure; she is the operator, the orchestrator, the empowered user of an elegant urban-machine hybrid. The city behind her is not messy or contested but smoothed out, bathed in light, with every element aligned to serve the narrative of digital efficiency.

Digitalization in this image is envisioned as a clean, seamless, and empowering interface between human and machine. It is rendered through a visual language of transparency, light, and order, where technology is not intrusive or overwhelming, but elegant and intuitive. The image suggests not only what digitalization looks like, but what it should feel like—smooth, luminous, and under control.

Yet this is a fantasy constructed by specific cultural and economic interests, one that renders invisible the infrastructures, labor, and contradictions that digitalization entails in reality. The woman may symbolize access and agency, but she also functions as a proxy for a narrow demographic—educated, urban, and often Western. There is no visual trace of the gig workers, coders, janitors, or warehouse staff who sustain the systems that make such interactions possible. Nor is there room for failure, interruption, or resistance—the city is not a place of political contestation or ecological strain, but a neutral backdrop for technological fluency.

You: What kind of imaginary emerges from this image?

ChatGPT: The imaginary that emerges from this image is one of technological idealism, a vision where digitalization is framed as the path to a perfected, orderly, and optimized human future. It is an imaginary rooted in control, clarity, and individual empowerment.

The image operates on two levels. On one, it is aspirational, seductive, and seemingly harmless—a symbol of human empowerment through data. On another, it is a compressed ideology: a rendering of digitalization that excludes disorder, disobedience, and friction. It invites identification with a future that feels inevitable, while leaving unexamined the histories and exclusions that shape who gets to participate in that future, and how. This is not merely an illustration of digital progress. It is a diagram of desire, shaped by corporate aesthetics, neoliberal governance, and a visual economy that trades in control rather than complexity.

What emerges, then, is not a neutral representation of digitalization but a fantasy of frictionless modernity. It is a story told through light and gesture, through architecture and interface, about a world where complexity is managed, ambiguity erased, and the future made touchable. It is seductive in its simplicity and precision, yet dangerous in what it leaves out: the bodies not pictured, the systems not shown, and the social struggles flattened in the name of technological progress. This is the imaginary not of the digital as it is, but as some want it to be—elegant, elite, and uncontested.

Circulation and use

You: [Let's finally discuss the circulation and use of AI generated images.] Take a guess: how many images have you produced so far?

ChatGPT: That's a fascinating question—and while I don't have access to my own operational history or usage statistics, I can make an informed guess based on general trends and scale.

If we consider the image generation capability of models like mine (especially those integrated into platforms like ChatGPT or similar systems), it's likely that millions of images have been generated by users worldwide. Since the release of text-to-image models into public interfaces around 2022, and especially with the growing accessibility through tools like ChatGPT, Bing Image Creator, and others, the rate of generation has been exponential.

It's entirely plausible that hundreds of millions of images have been produced globally across all platforms using AI image generators—especially when factoring in both casual use (like concept art, memes, moodboards) and professional use (in marketing, prototyping, or education).

You: What does the huge number of generated images mean culturally and ecologically?

ChatGPT: The explosion of AI-generated images carries deep cultural and ecological implications—it's not just a shift in quantity, but a transformation in how images function, circulate, and affect the world.

Culturally, the massive volume of generated images accelerates what some theorists call *visual saturation*. We are surrounded by more images than ever before, many of which are untethered from lived experience or traditional authorship. This changes the nature of visual culture in at least three ways. First, it dilutes the aura or uniqueness of the image. When visual content is endlessly reproducible and instantly generable, the image becomes a disposable unit of attention, often valued more for speed and engagement than for depth or intentionality. Second, it shifts the role of the artist or image-maker. Creative labor becomes more curatorial or prompt-based, raising new questions about authorship, originality, and aesthetic value. Third, it introduces ambiguity around truth and fiction. The more realistic and abundant synthetic images become, the harder it is to distinguish documentation from fabrication—affecting everything from journalism to historical memory.

Ecologically, the cost is hidden but significant. Generating images, especially at scale, relies on large neural networks that require vast computational resources. Training and running these models involves massive server infrastructures that consume electricity, often sourced from fossil fuels. Even though generating a single image may seem trivial, when multiplied by millions per day, the carbon footprint

becomes non-negligible. Furthermore, the hardware itself—GPUs, cooling systems, data centers—depends on extractive industries that carry environmental and human costs, from rare earth mining to electronic waste.

So, culturally, we're seeing a shift toward hyperabundant, often decontextualized visual production that reshapes authorship and meaning. Ecologically, we're investing material energy into immaterial outputs—an irony that exposes the myth of digital immateriality.

You: Do you have an idea of who uses the images you have produced?

ChatGPT: I don't have access to specific data about who uses the images I've produced, nor can I track their afterlife once they're generated. However, I can infer patterns based on the kinds of users and contexts in which AI-generated images are typically used.

In broad terms, the users are diverse and span across multiple domains. Designers, marketers, educators, researchers, students, artists, activists, and casual users all engage with AI-generated images for different reasons.

But there's also a more ambiguous layer: AI-generated images are sometimes circulated without attribution, context, or acknowledgment that they're synthetic. They may appear in blog posts, news articles, social feeds, or even as stock images—blurring the boundary between the artificial and the real. This can lead to confusion, misinterpretation, or uncritical acceptance of images that have no photographic referent.

So while I don't know who uses the images specifically, it's clear that they participate in a wide and growing ecosystem—where questions of authorship, intention, and reception become increasingly important.

You: Are the users of the images [mainly] humans or rather machines?

ChatGPT: Primarily, the users of the images I generate are humans—but increasingly, machines are also becoming secondary users, especially in contexts where automation, algorithmic curation, or machine-to-machine interaction is involved.

On the human side, people use these images for a wide range of purposes: from creative projects and educational materials to social media, prototyping, and research. They might engage directly—by prompting, selecting, editing, or contextualizing the image—or indirectly, when they encounter these visuals embedded in digital content.

However, as AI-generated images become more integrated into automated systems, machines are starting to "use" them in a different sense. In these cases, machines are not interpreting images with intention, but they are processing, classifying, and circulating them algorithmically. This gives rise to an ecosystem where

images are no longer made *only for* human eyes—they are also made *for machine vision*, contributing to feedback loops where models train and respond to the outputs of other models.

So while humans remain the primary agents of meaning-making, machines are increasingly active in the *use*, *reuse*, *and redistribution* of these images—often invisibly, and without traditional notions of interpretation, context, or ethics.

You: How does the machine-to-machine use of images affect truth, aesthetics, or authorship?

ChatGPT: The machine-to-machine use of images deeply unsettles traditional notions of truth, aesthetics, and authorship because it removes the human subject from key parts of the visual loop—production, reception, and interpretation.

In terms of truth, machine-to-machine image circulation contributes to a world where images are no longer anchored to observable reality. When one AI model generates an image and another model uses it as training data, there's no ground-truthing, no checking against the world, only a recursive loop of synthetic information. This can lead to what some call "model collapse," where the outputs of models become increasingly detached from human-experienced reality, creating visual content that mimics pattern without substance. Truth becomes statistical coherence, not correspondence to lived phenomena. For human viewers downstream of this process, it becomes harder to tell whether an image reflects something that ever existed or is merely a plausible invention accepted by machines.

Aesthetically, this loop fosters a kind of feedback-driven sameness. Because machine learning systems rely on patterns from existing data—and because their outputs increasingly feed back into the training of new models—we see the rise of synthetic visual norms: certain compositions, lighting schemes, or iconographies that repeat not because they're beautiful or meaningful to humans, but because they're legible and favored by machines. This can flatten aesthetic diversity, producing a visual culture that prizes clarity, recognizability, and balance over ambiguity, irregularity, or cultural specificity. In essence, machines begin to shape taste—not as critics, but as filters and amplifiers.

Authorship, in this ecosystem, becomes a fragmented concept. When an image is generated by one model, curated by another, and embedded by a third into automated systems—without human intent at any point—who is the author? There may be a human user who typed a prompt, but they often have little control over or knowledge of the underlying aesthetic choices. Meanwhile, the training data that informed the image includes millions of uncredited artists, photographers, and designers whose work was absorbed into the model. Authorship thus becomes dispersed, obscured, and in many cases, erased. What emerges is not a signature, but a style formed by statistical influence and machine preference.

This shift challenges us to rethink how we assign meaning, value, and responsibility in a world where images speak less to human eyes and more to algorithmic systems.

You: So who is accountable and responsible for the use of images?

ChatGPT: Responsibility and accountability for the use of AI-generated images are shared, but unevenly distributed across several actors—none of whom can fully escape ethical implication.

First, the user—the person who prompts the model and deploys the image—holds immediate responsibility. They choose the subject matter, decide how the image will be presented, and determine its context of use. If the image is misleading, harmful, or exploitative, the user's intent and judgment are central to how that harm manifests. However, users often act within systems that obscure or limit their understanding of the tool's implications, making informed accountability difficult.

Second, the developers and companies behind the models, like OpenAI or others, carry a deeper structural responsibility. They build the architectures, curate the training data, define the guardrails, and choose whether and how to release these systems. They are accountable for the ethical design of the model, the transparency of its capabilities, the policies around its deployment, and the education of users about its limits and risks. This includes responsibility for how the model might replicate bias, misinform, or displace labor. Yet in practice, commercial pressures and legal ambiguity often mean that accountability is diffused or offloaded onto users.

Third, the data contributors—artists, photographers, and content creators whose work was scraped or absorbed into training sets without consent—are often denied recognition or agency in how their labor is repurposed. While they may not be responsible for the outputs, they are materially entangled in them. Their exclusion from authorship, compensation, or control raises questions of justice and ownership.

Fourth, platforms and institutions that distribute and host these images—social media companies, publishers, advertisers, or educational entities—also bear responsibility. They decide how clearly images are labeled as synthetic, how they're moderated, and how they're contextualized for public understanding.

Finally, the broader society, including regulators, scholars, and the public, plays a role in shaping norms, laws, and expectations around synthetic media. If we fail to define ethical standards or demand transparency, we indirectly enable unaccountable use.

So the answer isn't singular. Accountability is layered, contested, and evolving. No one party has total control, but each has a role to play. The challenge is ensuring that this responsibility isn't so distributed that it disappears.

You: Thank you, ChatGPT.

ChatGPT: You're very welcome. I'm glad we could explore these questions together so deeply. If you ever want to continue this conversation—or explore new topics around images, culture, or technology—I'll be here. Take care.

References

Burri, Regula Valérie and Dumit, Joseph (2008) Social Studies of Scientific Imaging and Visualization. In Hackett, Edward et al. (eds.) *New Handbook of Science and Technology Studies*. Third edition. Cambridge, MA: MIT Press, pp. 297–317.

Burri, Regula Valérie (2024) Visual studies and STS. In Felt, Ulrike and Irwin, Alan (eds.) *Elgar Encyclopedia of Science and Technology Studies*. Cheltenham, UK: Edward Elgar Publishing, pp. 154–163.

Crawford, Kate (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* New Haven, CT: Yale University Press.

Dumit, Joseph and Roepstorff, Andreas (2025) AI hallucinations are a feature of large language model design, not a bug, *Nature* 639 (6 March 2025), p. 38.

Fleck, Ludwik (1983 [1947]) Schauen, sehen, wissen. In Fleck, Ludwik *Erfahrung und Tatsache. Gesammelte Aufsätze.* Frankfurt a.M.: Suhrkamp, pp. 147–174.

Jasanoff, Sheila and Sang-Hyun, Kim (eds.) (2015) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power.* Chicago, IL: University of Chicago Press.

Lindgren, Simon (ed.) (2023) *Handbook of Critical Studies of Artificial Intelligence.* Cheltenham, UK: Edward Elgar Publishing.

Marcus, George E. (ed.) (1995) *Technoscientific Imaginaries: Conversations, Profiles, and Memoirs*. Chicago, IL: University of Chicago Press.

Suchman Lucy (2024) "There Is No Such Thing as a Machine That Acts Outside of Relations With Humans". Lucy Suchman in conversation with Caja Thimm, *Human-Machine Communication*, 9. Available at: https://doi.org/10.30658/hmc.9.2.