IMAGE
Zeitschrift für interdisziplinäre Bildwissenschaft
Ausgabe Nr. 42 (2/2025), S. 155-168
ISSN 1614-0885
DOI: 10.1453/1614-0885-2-2025-16663

HERBERT VON HALEM VERLAG

Charlotte Brachtendorf

Possible Fashion Images: Operative Ekphrasis and the Reduction of Fashion Through Multimodal AI

Abstract

With the rise of multimodal AI tools, such as DALL-E, Midjourney, or Stable Diffusion, it has never been easier and faster to produce compelling fashion images. While the proliferation of AI-generated fashion images can be seen as the ultimate triumph of the fashion image over both fashion text and fashionable garments, this paper critically examines how these images are produced, what exactly they show, and ultimately, how this impacts our understanding of fashion. Building on current research on multimodal AI, specifically text-to-image models, I argue that AI-generated fashion images are produced through textual means, that is, through "operative ekphrasis" (BAJOHR 2024: 83). This means that the generation of fashion images created via multimodal AI is based on textual prompts that extract statistically probable images from a latent space, which, in turn, is the result of previously indexed and labeled image-text pairs. Because the indexing of images requires a reduction of complexity, images generated through multimodal AI are often generic. In the case of fashion images, this is reflected in the lack of texture and the reproduction of fashion photography's conventions. As fashion photography is not necessarily a representation of 'the world' or even fashionable garment*s*, it resembles AI-generated images more than other photographic genres. AI-generated fashion images are then not images of fashionable *garments*, but images about fashion *images*—a statistically probable permutation of fashion photographs from the past (cf. MEYER 2023: 108). As such, AI-generated fashion images are not only the product of a mediated collective imaginary but also feed back into it. The view on fashion shaped by AI-generated fashion images is then based on reduction and stylized normativity.

Dank multimodaler KI-Tools wie DALL-E, Midjourney oder Stable Diffusion war es nie einfacher ansprechende Modebilder zu produzieren. Obwohl die Verbreitung KI-generierter Modebilder als finaler Triumph des Modebildes über den Modetext sowie die Modekleidung gelten könnte, zielt dieser Artikel auf eine kritische Betrachtung des Phänomens, indem er in den Blick nimmt wie diese Bilder produziert werden, was genau sie zeigen und letztlich welches Verständnis von Mode daraus folgt. Vor dem Hintergrund aktueller, medienwissenschaftlicher Forschung zu multimodaler KI und speziell Text-zu-Bild Generatoren argumentiert dieser Artikel, dass KI-generierte Modebilder durch „operative Ekphrasis" (BAJOHR 2024: 83), also durch Text produziert werden. Die Generation dieser Modebilder basiert of textuellen ‚Prompts', die ein statistisch wahrscheinliches Bild aus einem ‚Latent Space' extrahieren, der seinerseits das Ergebnis verschlagworteter Bild-Text-Paare ist. Da die Verschlagwortung von Bildern auf einer Reduktion von Komplexität beruht, zeichnen sich KI-generierte Bilder häufig dadurch aus, dass sie generisch sind. Im Kontext von Modebildern äußert sich dies vor allem in der Abwesenheit von Texturen sowie in der Reproduktion modefotografischer Konventionen. Modefotografie bedeutet dabei nicht zwangsläufig die Abbildung von ‚Welt' oder Modekleidung und steht daher KI-generierten Bildern näher als andere fotografische Genres. Was KI-generierte Modebilder daher zeigen ist keine Mode(kleidung). Stattdessen zeigen sie Bilder, die von Modebildern handeln; sie sind eine statistisch wahrscheinliche Permutation von Modefotografien aus der Vergangenheit. KI-generierte Modebilder sind aber nicht nur das Ergebnis einer medialisierten kollektiven Imagination, sondern speisen sich auch in eben diese ein. Daher entwerfen und verstetigen KI-generierte Modebilder ein stilisiertes, normatives und letztlich verkürztes Bild von Mode.

Introduction: AI-generated Fashion Images

In the spring of 2023, just a few months after text-to-image AI was made publicly available, an image went viral, showing Pope Francis in an oversized Balenciaga puffer jacket. What has since become known as the "Balenciaga Pope" meme (MEYER 2024: 39) is also one of the most prominent intersections of fashion and multimodal AI to date. Rather than an actual photograph of Pope Francis dressed in Balenciaga, this image was generated with Midjourney by a Chicago construction worker (cf. WILDE 2023: n. p.). Lukas Wilde (2023: n.p.), who has traced the spread of the image through Reddit and Twitter, notes that many viewers were initially fooled, despite some pointing out the visual inconsistencies early on. Soon enough, viewers engaged in "wild forensis", the practice of detecting

AI-generated images through incongruent details (MEYER 2024: 39). That this viral AI-generated image has placed a fashionable garment front and center seems more coincidental than deliberate, yet it points to the potential of multimodal AI for the production of fashion images.

Fashion magazine covers seem to lend themselves particularly well to AI-generated images. *Cosmopolitan*, for example, published its first cover generated with OpenAI's DALL-E 2 in 2021 (cf. LIU 2021; see also MEYER 2023). The cover depicts a female astronaut in a space suit stomping across an extraterrestrial desert, bathed in magenta light. While the absence of a face breaks with the conventions of fashion magazine covers, the issue highlights the fact that its cover is generated with AI. Claiming "it took only 20 seconds to make", *Cosmopolitan* highlights the speed at which multimodal AI can compute input, while simultaneously concealing the tinkering on the part of the user to achieve the desired result (LIU 2021: n. p.; MERZMENSCH 2024: 41). *Glamour Bulgaria* pulled a similar stunt in 2023, with even more realistic results. An AI-generated image of the former Miss Virgin Islands, Lisa Opie, graced the cover, accompanied by a full fashion spread of Opie as a "futuristic Barbie" inside the issue (GUZMAN 2023: n. p.). These images were created by a woman named Fran, who runs an Instagram account featuring AI-generated fashion images, boasting nearly 100,000 followers (cf. @ai_fashion_photos on Instagram; DENINNO 2023). On this account, Fran regularly publishes so-called carousels, posts that feature up to 20 thematically similar images, often marked as "photoshoots" or "editorials" (cf. @ai_fashion_photos on Instagram, posts from 13.02.2025 and 13.12.2023). Aside from her activities on Instagram, Fran was also a participant in the first AI Fashion Week, held in 2023. It was organized as a contest and yielded more than 400 submissions of AI-generated fashion images, leading to a second season of the fashion week in 2024 (cf. AI FASHION WEEK n.d.). Participants were asked to submit a total of 30 images, consisting mainly of catwalk looks but also behind-the-scenes or front-row images (cf. ibid.).

This brief overview of the successes AI-generated fashion images have had in recent years suggests that AI image generation has been welcomed with open arms by the fashion industry, not least because they are easier and faster to produce. In line with the observation and approach of this special issue, fashion appears to have become entirely pictorial. Yet, such heralded innovations often warrant a critical investigation. As such, this paper aims to problematize AI-generated fashion images in two ways. First, by examining how these images are produced and second, by asking what consequences the mode of production has for fashion and its images. To address the first point, I draw on current research on multimodal AI in the field of media studies. Although the output of multimodal AI comes in the form of images, they are produced through textual prompts, which can be described as an operative ekphrasis, where textual

prompts elicit a statistically possible image from a latent space (cf. BAJOHR 2024: 83; BOLWIN 2024: 253). With this textual mode of image production in mind, I turn to the ramifications of AI image generation, specifically examining what AI-generated fashion images show. Certainly, AI-generated fashion images are not images of fashionable *garments*, even if 'conventional' fashion photography lends itself well to imitation through multimodal AI. Instead, drawing on Roland Meyer (2023: 108), I argue that they show images about fashion *images*; that is, a statistical permutation of fashion photography's conventions, which leads to a reductive view of fashion that (re-)establishes normative ideas of what and who is fashionable.

1. AI-generated Fashion Images as Operative Ekphrasis

Although Artificial Intelligence (AI) is widely considered a marketing term aimed at naturalizing and anthropomorphizing stochastic Machine Learning (ML), it nonetheless "allows for a linguistic pragmatism" in the burgeoning field of Critical AI Studies, where AI is understood as "an assemblage of technological arrangements and sociotechnical practices, as concept, ideology and *dispositif*" (RALEY/RHEE 2023: 188, original emphasis). This article focuses on so-called multimodal AI, particularly text-to-image generators, which can process various types of data, including images *and* text (cf. BAJOHR 2024: 85). These models function as so-called "artificial neural networks"—another suggestive term intended to conjure the mechanisms of the human brain (ibid.: 80). They are trained on vast amounts of text-image pairs so that they can generate an image out of a natural language description known as a "prompt" (MERZMENSCH 2024: 31). The desire to steer and fine-tune the visual output through prompting with precise or poetic language has since become known as "promptology" or "prompt-engineering," both terms that seem to imply a scientific basis (BOLWIN 2024: 259).

As others have noted, the interplay of text and image is particularly interesting here: "Even the first experiences with DALL-E point to a paradigm change in the genesis of images. If images were previously created through craftsmanship (either with oil paint, cameras or Photoshop), they are now formed through the medium of language" (MERZMENSCH 2024: 37, my translation). DALL-E and similar models can perform this transformation from text to image because they can "create complex semantic connections" (ibid.: 34, my emphasis) between textual descriptions and their visual counterpart, even if the *semantics* here are artificial and therefore limited (cf. BAJOHR 2024: 86). For Bajohr, this marks a departure from what he calls a "sequentialist paradigm" of computation forming the basis of classic algorithms where step-by-step instructions lead to a deterministic

result (ibid.: 80). In contrast, the "connectionist paradigm" underpinning multimodal AI is governed by a stochastic logic: Because the semantic connections 'learned' through the training data are "embedded in the [artificial neural] network's structure", the visual output of these models is a *probable* result of a given prompt (ibid.: 82). It is precisely this emergence of semantics that leads Bajohr to describe the processes of multimodal AI as "operative ekphrasis" (ibid.: 83).

The concept of "ekphrasis" has been discussed in various academic disciplines and is the subject of countless publications. For this context, it suffices to note that the term originally referred to the "practice of evoking images by means of verbal, often poetic expression" in ancient Greece, but was later co-opted as a concept and method by art history (BOLWIN 2024: 255). As media scholar Charlotte Bolwin aptly points out, it has evolved "from an art of description to a description of art" (ibid.). Ultimately, however, Bolwin aims for a broader understanding of ekphrasis, one that describes "the relation of language to visual culture in *toto*" (ibid., original emphasis). In his transfer of ekphrasis to multimodal AI, Bajohr defines it as a *"computational operation that correlates text and image"* (BAJOHR 2024: 83, original emphasis). In this semantic correlation of image and text, Bajohr argues that multimodal AI effectively deconstructs the image-text divide (cf. ibid.: 84). It is in "the shared meaning in the artificial neuron" that Bajohr locates the "structural identity between image and text" (ibid.: 86).

Yet, the observation that multimodal AI performs an operational ekphrasis applies to all AI-generated images. To understand what this means in the context of fashion, I turn to Barthes' distinction between 'written clothing' and 'image-clothing' (BARTHES 1990: 3) and highlight how AI-generated fashion images challenge this analytic paradigm. As one of the first scholars who took fashion magazines seriously as a topic of scientific analysis, Barthes built on Saussurean semiology to devise a structuralist approach to fashion. Within fashion magazines, Barthes deemed 'written clothing' the most appropriate for analysis because it "has no practical or aesthetic function: it is entirely constituted with a view to a signification: if the magazine describes a certain article of clothing verbally, it does so solely to convey a message whose content is: *Fashion*" (ibid.: 8, original emphasis). From this, Barthes concluded that the verbal codes of fashion magazines produce fashion as a system of signification (cf. ibid.: 51).

Returning to the notion of ekphrasis, the concept of image-text transduction is not entirely new to fashion representation, even if it has not been explicitly framed as such. According to Barthes, "written clothing" and "image-clothing" both refer to a third structure, that is "real clothing" (ibid.: 4). These three structurally disjunct garments can be transformed into one another through so-called "shifters" (cf. ibid.: 5-7). For example, Barthes identifies sewing patterns and photography as shifters that can translate the technological structure of a real garment to the iconic structure of an image (cf. ibid.: 6). Notably,

Barthes does not devise a shifter for the transformation of 'written clothing' into 'image-clothing', only the other way around. Although this transformation could be described as ekphrastic in the sense that it addresses the relationship between textuality and visuality, it also differs significantly from the operative ekphrasis performed by multimodal AI: Where 'written clothing' and 'image-clothing' are usually perceived simultaneously (if not in *The Fashion System* itself, then in the magazines Barthes analyzed), the temporal structure of AI-generated fashion images is asynchronous. The images produced through generative AI are an effect of the textual prompt, appearing only after the prompt has been issued. Similarly, AI-generated fashion images turn the relationship between iconic and verbal structures on its head. In fashion magazines, Barthes observed that the translation of image into text is often achieved through elliptical sentences. But rather than omitting details, the prompts that produce the AI images are often lengthy and meticulous descriptions of images not yet visible to a spectator.

Fig. 1: AI-generated fashion images, created with Midjourney, Instagram post by @ai_fashion_photos, 13.02.2025.

This highlights one of the most intriguing aspects of AI-generated fashion images. While the images in fashion magazines are concrete and indexical, those generated through multimodal AI are statistically possible. Despite the importance placed on 'prompt-engineering', multimodal AI produces unpredictable and unrepeatable results (cf. MEYER 2023: 102). This can be observed in the "photoshoots" posted on the Instagram account @ai_fashion_photos, where the images grouped together in carousel posts are often permutations of one and the same motif. In a post from 13 February 2025, for example, all images show a white, blonde, and slim model, dressed in a pastel gown, posing in or in front of a pink car in Utah's salt flats (see Fig. 1). Some of these images are so similar that it is reasonable to assume they were generated through the same prompt. And yet, they are not exactly the same. Roland Meyer therefore conceptualized

AI-generated images as "a statistical possibility in a latent image space fed by images from the past" (ibid.). In the context of AI image generation, latent space refers to the entirety of possible images, clustered in correspondence to semantic concepts (cf. ibid.: 106). In this logic, prompts are more like "search queries" within this latent space, locating and thus producing images through their description (ibid.: 102–103).

2. What Do AI-generated Fashion Images Show?

Where does this latent space of possible images come from, then? By now, it is common knowledge that the data used to train AI tools has a profound influence on the results it can produce. For example, if the training data contains gendered or racial biases, these will be reproduced in generative AI (cf. OFFERT/PHAN 2022). While such biases also apply to multimodal AI, text-to-image tools face another challenge: The images used to train the tools need to be assigned labels, that is, textual correspondences, for the artificial neural network to 'learn' their meanings. Meyer calls this "semantic pre-processing", which includes adding metadata, labels, indexes and tags to the training images (MEYER 2023: 105–106). Only when the training images are paired with textual labels can they be used to train the semantic connections within the latent space of multimodal AI (cf. ibid.). Consequently, both Bolwin and Meyer stress the role of labeling as an "infrastructural precondition" of multimodal AI; a necessary step before users can start prompting (MEYER 2023: 105, see also BOLWIN 2024: 258–261).

However, labeling, indexing, and tagging images are by no means unequivocal processes. Images are often ambiguous and indeterminable, allowing for multiple possible meanings (cf. BOLWIN 2024: 260). Yet, the process of training multimodal AI requires the elimination of this ambiguity and diversity, so that the data can be used for the computation of probabilities (cf. ibid.). Therefore, the labeling process seeks to normatively assign "exact correspondences" to the training images at hand (ibid.). Labels are usually assigned by precariously employed "ghost work[ers]" (MEYER 2025: 6) or by algorithms that crawl the internet and ascribe images with textual information previously associated with similar visual patterns (cf. BOLWIN 2024: 261). In any case, this 'semantic pre-processing', based on a reduction of complexity, leads to images that are "not only generated, but also generic" (ibid.). Contrasting AI-generated images to photographic images, Meyer elaborates on their genericness: "Instead of showing you, for example, this sunny afternoon today in this specific park here, it shows you a generic park that existed nowhere and never" (MEYER 2025: 7, original emphasis). This is because the processes of labeling and indexing only 'teach' the artificial neural network abstract concepts. As such, multimodal AI cannot produce an

image of a specific garment, a coat, for example; it can only generate an image derived from the concept of 'coatness', combined with other concepts, it has learned through the previously labeled training images (cf. ibid.)

Fig. 2: AI-generated fashion images by Gianluca Traina, Christina Berezhneva, and Matilde Mariano, created with Midjourney and Photoshop, AI Fashion Week 2023 and 2024.

Examining the images submitted to AI Fashion Week, this genericness is evident in two aspects. First, many images appear exceedingly smooth. Although the images generally appear photorealistic and the close-ups often reveal details, the depicted surfaces exhibit barely any texture. The garments have no wrinkles, except for the desired drape; models have poreless skin, and the setting often recedes into a fuzzy background (see Fig. 2). Instead, they display visual patterns based on the *concepts* of drape, skin and background. This highlights that these images are not the result of registered "light nor brushstrokes" (ERVIK 2023: 53), but stochastic interpolations of possible data points. And secondly, the images presented at AI Fashion Week are generic in that they are characterized by a "visual conventionality" (MEYER 2023: 105). They adhere closely to established visual codes of fashion week reporting. For instance, the (imaginary) photographer is positioned at the end of the catwalk, seemingly capturing the models mid-walk, and the blurry background often alludes to the fashion show audience seated along the catwalk.

Interestingly, AI Fashion Week complements the 'traditional' catwalk images with those that show the front row or the backstage (see Fig. 3). Although such 'candid' images are ingrained in the visual repertoire of fashion week coverage and can therefore be considered fashion images in their own right, their inclusion in AI Fashion Week is peculiar. Clearly, the fashion shows envisioned for AI Fashion Week have never materialized as real-life events. Still, behind-the-scenes

images, which usually document and often stage the production of a fashion show, present themselves as quasi-evidence of a show that has physically taken place, and not as one that was merely conjured through the computation of multimodal AI. They provide an imaginary backstage to an (equally imaginary) catwalk, painting a more fleshed-out picture of the fictional fashion show. While these images can be seen as a strategy to heighten the realism of AI Fashion Week, they end up highlighting the factitiousness of these fashion shows all the more. As the name already indicates, viewers likely know that the images in question are AI-generated. Rather than authenticating these images, the insistence on their 'candid', seemingly documentary nature raises a pressing question: If we are not seeing the documentation of a fashion show and its backstage, what exactly do AI-generated fashion images show?

Fig. 3: AI-generated behind-the-scenes images from AI Fashion Week 2023 by Anya Klyueva and Gaby Roses, created with Midjourney..

Andreas Ervik warns of "looking *at* the image as a photograph", arguing that it is necessary to take "into account how the image-producing technology mediates the viewer's relation to what is depicted" (ERVIK 2023: 45, original emphasis). Indeed, the images of AI Fashion Week are not necessarily about fashionable clothes, but can be read as a meta-commentary on fashion (images). As Roland Meyer insists, any given AI-generated image is not a depiction of reality, but rather one possible permutation of an immense number of image-text pairs (cf. MEYER 2023: 108). Consequently, they are "not images of the world, but images of images—indeed, ultimately images about images, filtered through language" (ibid.). In the context at hand, this means that AI-generated fashion images are not images of fashion, but images about fashion images produced through textual means. Instead of three-dimensional fashionable garments, they show a

stochastic result of fashion photography's genres (e.g., runway photography or editorial).

AI-generated fashion images still appear to us as fashion photographs, and often require a second look to discern their genesis, precisely because fashion photography relies on performatively established conventions. Even 'regular' fashion photographs have never been understood as a representation of a reality known as 'world'. Barthes himself noted as much, in the appendix to *The Fashion System:* Fashion photography, he argues, turns the world into "a decor, a background or a scene, in short, a theater" (BARTHES 1990: 300). It empties the signified of its connection to reality: "The world, *everything which is not the garment*, is exorcised, rid of all naturalism" (ibid.: 303). This is in stark contrast to Barthes' other writing on photography, where he attests the photographic image a necessary connection to a real-world referent (cf. BARTHES 1981: 76). Fashion photography is then unlike other types of photography in that it does not show "that-has-been" (ibid.: 77; see also BARTHES 1990: 4) but imagines what could—or should—be. In this regard, fashion photography has a closer affinity to AI-generated images than to other types of photography. If fashion photography is already a scene removed from reality itself, this genre is easily translatable into AI-generated images.

While Barthes maintains that fashion photography makes everything except the depicted garments implausible (cf. BARTHES 1990: 300), even the fashionable garments themselves can be brought into question. For instance, the philosopher Gilles Lipovetsky observes that contemporary fashion photography is less inclined to depict garments—they can be entirely absent—and more interested in "creating an atmosphere" (LIPOVETSKY 2002: T8). Consequently, fashion photography does not necessarily uphold an indexical relationship between a material object and its photographic representation. In fact, as Susan Sontag argues, the photographic image has already taken precedence over the material fashion object. In a short essay published in Vogue, she argues that fashion "has become something that is almost entirely visual—that is, photographic. As fashion becomes pure appearance, it finds its perfect summing-up in photographs. [...] More and more, fashion is fashion photography" (SONTAG 1978: 508, original emphasis).

Certainly, Sontag's argument can be read as a criticism of postmodern visual culture, but more interestingly, her and Lipovetsky's understanding of fashion photography lends itself to discussing the relationship between AI-generated fashion images and the fashionable garments depicted in them. Whereas Barthes contended that the iconic structure of the fashion photograph did not show a real garment but, at the very least, referred to its existence (cf. BARTHES 1990: 4–5), this is not the case for AI-generated fashion images. Although they follow the conventions of fashion photography and seemingly depict fashionable garments,

these garments exist only in the image. Consequently, it is impossible to speak of an indexical relationship between image and garment, simply because there is no real-life garment the image could be referring to. If anything, the garment has become indistinguishable from the image that depicts it. Therefore, AI-generated fashion images appear as fashion photographs in Barthes' sense, but because they do not establish an indexical link between a real object and its depiction, they are more in line with Sontag's and Lipovetsky's understanding of fashion photography. Fashion photography is thus structured in such a way that both 'the world' and the fashionable garments themselves are no certainty. If fashion photography's removal from 'the world' has already positioned the genre as a prime candidate for AI image generation, the absence of an indexical relationship between image and garment heightens this affinity to AI-images. Thus, I contend that fashion photography is not just a suitable but an ideal subject for the simulation through multimodal AI.

Despite its convenience, it is not necessarily a desirable development that fashion photography can be easily imitated through multimodal AI. Considering that multimodal AI is trained on vast amounts of previous image-text pairs, any AI-generated image is nothing but a possible rearrangement of past images. For fashion images, a type of image that operates on the assertion of its newness, this perpetual self-referentiality can pose a problem. After all, how can these images be understood as new, when they are really a statistical permutation of "images from the past" (MEYER 2023: 102)?

However, even beyond the issue of novelty, multimodal AI represents a pre- and refiguration of what we traditionally think of as fashion. As Andreas Ervik argues, "AI image generators are an acquisition and extension of the *collective imaginary*" (ERVIK 2023: 52, original emphasis). Ervik is less interested in claiming that multimodal AI possesses the same capacities for imagination as humans and thus distances himself from the anthropomorphization of AI (cf. ibid.). Instead, he suggests that AI-generated images are the result of a "machinic imaginary" (ibid.: 54) produced by a mediatized collective imaginary. As AI-generated images are used as additional training data for multimodal AI, they also feed back into this imaginary, where they become the basis for further image generation. Simultaneously produced by and productive of this machinic imaginary, AI-generated images inscribe conceptions (over-)simplified through semantic indexing into our collective imagination. For example, LAION 5B, a database of training images, was compiled from images predicted to be visually pleasing to viewers. However, for this database, the judgements on which images are perceived as pleasurable come from a relatively small group of European and American men (cf. MEYER 2025: 7–8). But beyond this example, multimodal AI evaluates all images it produces based on how users interact with them, whether they are upscaled, downloaded, shared, or discarded that loop back

into the types of images produced to begin with (cf. ibid.: 8). The result of these evaluation processes is what Meyer calls "a recursive algorithmization of taste: machines predict the ratings of human users, which are then used to optimize other machines to please the aesthetic expectations of other human users" (ibid.). It is then hardly surprising that the reductive labeling of training images paired with the optimization towards particular demographics reinforce each other to produce normative, "structurally conservative" (ibid.: 11) images. For fashion images, the reduction of complexity in semantic labeling, the reliance on past fashion photographs and the algorithmic evaluation on what is considered aesthetically pleasing results in a highly stylized vision of fashion that leaves little room for ideas not adhering to current or past fashionable ideals. Where many fashion practitioners today see fashion as a mode of creative expression that can push boundaries, such views tend to be left out of the equation in the production of AI-generated fashion images. Instead, these images are often less diverse and innovative in terms of fashion design as well as categories such as race, gender and ableism, amongst others. Any broadening in an understanding of who and what can be fashionable is ultimately reversed in multimodal AI. Therefore, the technologically mediated vision of fashion produced through AI-generated fashion images is an impoverished one. As such, AI-generated fashion images represent not only a novel way of producing fashion images but also stand to reconfigure and replete conceptions of fashion as they persist within the collective imaginary.

Conclusion

As AI-generated fashion images present themselves as a new and easier-to-produce type of fashion image, they reinforce a pictorial understanding of fashion. In light of the proliferation of AI-generated fashion images, this paper investigated how these images are produced, and how their mode of production discerns them from 'regular' fashion images. Drawing on current research on multimodal AI from the field of media studies, I argued that AI-generated fashion images are produced through textual, rather than pictorial means. Not only does multimodal AI rely on image creation through textual prompts, but it also harnesses semantic connections between image elements and their corresponding textual descriptions. AI image generation can therefore be described as an operative ekphrasis, where textual prompts localize and actualize possible images within a latent space. Unlike 'conventional' fashion images, the actualized AI-generated fashion images are then not concrete or indexical but a statistically possible and unpredictable result of a textual prompt.

Yet, before multimodal AI can be utilized to generate images, it is first trained on large quantities of text-image pairs, which result from the labeling and indexing of training images and produce the latent space in the first place. The process of indexing is invariably one of reduction, leading to fashion images that are generic both in their lack of texture and in their adherence to conventions of fashion photography. As such, AI-generated fashion images do not show fashion, or document fashion shows; they are a statistical re-interpretation of fashion photography. Similarly, fashion photography does not depict reality, and sometimes not even garments, making it a photographic genre that shares more with AI-generated images than with documentary photography. However, the concept of fashion created through AI-generated fashion images is a reductive one that is produced by semantic indexing and instantaneously reintegrates back into our collective imaginary. Through the lens of multimodal AI, fashion becomes increasingly stylized and normative. Consequently, the euphoria surrounding AI-generated fashion images must face its limitations: Not only are these images the result of text-based image generation, but more importantly, this algorithmically mediated mode of production repletes innovative and diverse views on fashion by (re-)establishing normative notions of fashion design, gender, and race.

Bibliography

AI FASHION WEEK: www.fashionweek.ai (15.06.2025)

BAJOHR, HANNES: Operative Ekphrasis. The Collapse of the Text/Image Distinction in Multimodal AI. In: *Word & Image* 40(2), 2024, pp. 77–90

BARTHES, ROLAND: *Camera Lucida. Reflections on Photography.* Trans. RICHARD HOWARD. New York [Hill and Wang] 1981

BARTHES, ROLAND: *The Fashion System.* Trans. Matthew Ward; Richard Howard. Berkeley [University of California Press] 1990

BOLWIN, CHARLOTTE: Digital Ekphrasis? On Language-Image-Relations in Contemporary AI's Imageries. In: *Studi di Estetica* 30, 2024, pp. 249–265

DENINNO, NADINE: First-ever AI Fashion Week Debuts in NYC. 'A New Realm of Creation'. In: *New York Post*, 20.04.2023. https://nypost.com/2023/04/20/first-ai-fashion-week-coming-to-nyc-new-realm-of-creation/?utm_source=url_sitebuttons&utm_medium=site+buttons&utm_campaign=site+buttons&fbclid=PAZXhobgNhZWoCMTEAAafzaqLDhsDoQKzAz1JPc6Gw2Zrq9SXg8ENovkDU3IoWubhDegMtwmZPS_U3OQ_aem_ZgUdoo8aF-qX69Hl_h7Jew (11.06.2025)

ERVIK, ANDREAS: Generative AI and the Collective Imaginary: The Technology-Guided Social Imagination in AI-Imagenesis. In: *Image 37(1)*, 2023, pp. 42–57

GUZMAN, ALYSSA: Glamour Bulgaria Cover Model Uses AI to Create Barbie-inspired Photoshoot. In: *New York Post,* 12.07.2023 https://nypost.

com/2023/07/12/model-lisa-opie-uses-ai-to-create-glamour-bulgaria-barbie-inspired-cover/?fbclid=PAZXhobgNhZWoCMTEAAae_MWHzXWXd74PNWlLorfFbt6d_v3iUCoqM81vA2AQJ4XRx13UmiqOBhs5hTg_aem_SWaFGI8H9eQke394LfcPhQ (1.06.2025)

LIPOVETSKY, GILLES: Modischer als die Mode. In: LEHMANN, ULRICH; JESSICA MORGAN (eds.): *Chic Clicks. Modefotografie zwischen Kunst und Auftrag.* Ostfildern-Ruit [Hatje Cantz] 2002, pp. T8–T11

LIU, GLORIA: The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover. In: *Cosmopolitan*, 21.06.2021. https://www.cosmopolitan.com/lifestyle/a40314356/dall-e-2-artificial-intelligence-cover/ (11.06.2025)

MERZMENSCH: *KI-Kunst*. 2nd ed. Berlin [Wagenbach] 2024

MEYER, ROLAND: The New Value of the Archive. AI Image Generation and the Visual Economy of 'Style'. In: *Image 37(1)*, 2023, pp. 100–111

MEYER, ROLAND: Spekulative Strategien. KI-Bilder, Memesis und wilde Forensis. In: *Fotogeschichte 172*, 2024, pp. 38–44

MEYER, ROLAND: „Platform Realism". AI Image Synthesis and the Rise of Generic Visual Content. In: *Transbordeur* 9, 2025, pp. 1–18

OFFERT, FABIAN; THAO PHAN: A Sign That Spells: DALL-E 2, Invisual Images and the Racial Politics of Feature Space. In: *arXiv:211.06323*, 26.10.2022. https://arxiv.org/abs/2211.06323 (13.06.2025)

RALEY, RITA; JENNIFER RHEE: Critical AI. A Field in Formation. In: *American Literature 95(2)*, 2023, pp. 185–204

SONTAG, SUSAN: The Avedon Eye. In: *Vogue*, September 1978, p. 461 and pp. 507–508

WILDE, LUKAS R.A.: AI-Bilder und Plattform-Memes: Post-digital, post-artifiziell, post-faktisch? In: *Kurzfilmfestival Tübinale*, 11.04.2023. https://www.tuebinale.de/aibilder-und-plattformen/ (29.07.2025)

Biography

Charlotte Brachtendorf is a fashion and media scholar. She is a research assistant for Cultural Studies in Fashion and Textiles at the University of Paderborn (Germany) and a member of the Research Training Group "Media Anthropology" at the Bauhaus-University Weimar (Germany), where she has recently submitted her PhD dissertation on the materiality of digital fashion. Her research focuses on the potential of media theory and media philosophy for the study of fashion.