Making up Data: Generative AI and the Epistemological and Moral Challenges of Synthetic Image Data

Making up Data

Generative AI and the Epistemological and Moral Challenges of Synthetic Image Data

A.K.M. Skarpelis, Assistant Professor, CUNY

Anna.skarpelis@qc.cuny.edu

Note : This is a first 7100 word short draft of the paper; I will have a more complete version ready by the fall.

The rise of generative AI has sparked troubling developments. Consider when Meta in January 2025 debuted a "sassy black queer" AI assistant called Liv, whose exposure as "digital blackface" saw the company quickly retire the character [1]. While the effort aimed to address racial bias, it unintentionally invoked the fraught history of racial transformation and deceit—whether through acts of passing, minstrelsy, or mockery. Most synthetic data is not problematic because it is artificial, but because their mode of creation echoes historically familiar forms of power and exploitation. This article takes a sociological lens to questions of knowledge and power around synthetic data. As a sociologist of science and technology, I study missteps with progressive intention as the ones mentioned above to try and articulate the special challenges that images pose for computational analysis and meaning making. In the course of this, I pull together the literatures on moral entrepreneurship in AI ethics [2], that on agency in machine learning [3], the STS literature on translation in the construction of scientific facts [4-6] and the economic sociology literature on performativity, especially where it concerns person categorization and classification [7-9]. This paper has two aims: first, to analyze synthetic data, especially image data, through an STS framework that categorizes it as part of a broader system of generated scientific facts; and second, to examine how power and ethics manifest in generative AI, using examples of progressive initiatives that unintentionally failed.

Keywords: artificial intelligence, ethnoracial ascription, generative AI, AI ethics, ChatGPT, inference, classification

1 INTRODUCTION: CONSIDER THE PIXEL

The generative turn in artificial intelligence has led to some sinister developments. Engineers developed and implemented technologies that led to the generation of a set of images of supposedly "German soldiers" from "1943" but included the likeness of an East Asian woman and a Black man; some years earlier, a group of Chinese researchers applied domain-to-domain translation, a hot new method in the field of computer vision, to ethnicize and racialize the faces of white people and turn them into people of color (see figures 1 and 2).

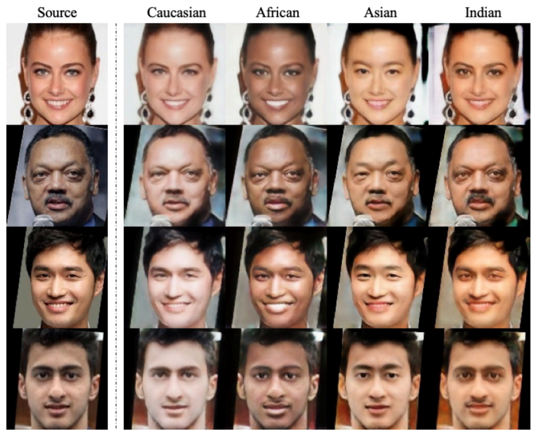

Figure 1. Image-to-image translation results of different racial domains. The first column shows source images and the other four columns are in turn converted to Caucasian, African, Asian and Indian

Figure 1. Racial "Translation", from Jianchen Ge et. Al [10], p. 1

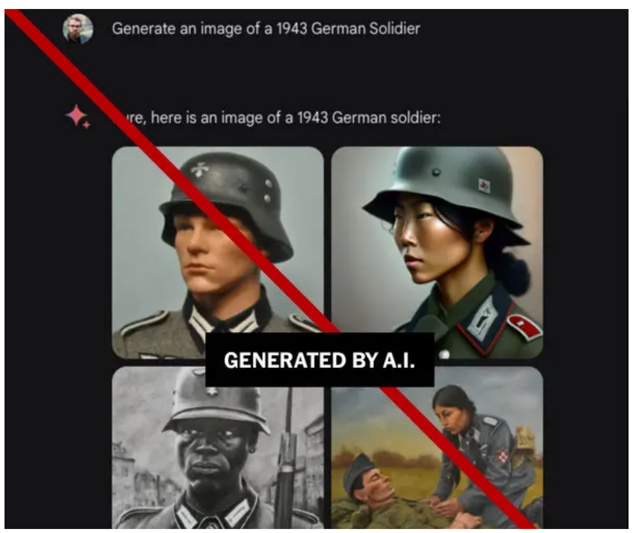

Figure 2. The infamous PoC Nazi, https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical

Both projects attempted to redress racial bias and neither intended for historical revisionism, nor harbored desires for “transracial”[1] refiguration. And yet, the images’ extraordinary clumsiness in a context of a troublesome history of acts of racial transformation – sometimes acts of passing, more often acts of mockery[2] – harbor the potential to teach us something more fundamental and generalizable about the relationship between the social and its visual representation in a world of generative artificial intelligence.[3] Connecting to FAccT interests in power, we build on the existing critical literature on generative AI [13-22] to outline three epistemological problems that arise out of these kinds of images in particular.

Some twenty years ago, David Foster Wallace wrote an essay called “Consider the Lobster” for Gourmet Magazine. The essay, whose title is indebted to a collection of essays by the greatest food writer of the 20th century, M. F. K. Fisher, made a splash because it thematizes the ethics of boiling a sentient creature alive [23, 24]. The morally dubitable process of lobster-boiling holds modern appeal, Wallace tells us, because the proximity to the event is extraordinarily close so that there is no “decomposition between harvesting and eating” [23: 56]. What makes this closeness alluring is not only the mundane aspect of food safety (contamination is less likely the less time passes between harvesting and consumption) but also the authenticity that derives from the proximity to the event – the testimonial value that accrues from witnessing the lobster alive and of having oversight over every step in the transition from life to death. By now, you are asking yourself: What does boiling lobsters alive have to do with pixels? [4]

Photography as a medium differs from painting in its proximity to the event. Henry Fox Talbot’s phrase the pencil of nature [27] poignantly illustrates that the process was considered to permit a more authentic and direct translation of what is in the world to a physical medium. There is less risk of distortion, less risk of the artists’ subjectivity barging in on the scene. Photography is also, compared to one-dimensional text, a spatial

[1] For the purposes of this piece, I do not engage with the question of whether “transracialism” is a thing or not. For an introduction to the controversy around the term, consult: [10] Mansi Hitesh. 2024. Troubling Transracialism: A Transnational Perspective. *Feminist theory* https://doi.org/ 10.1177/14647001241231717, [11] Ann Morning. 2017. Race and Rachel Doležal. *Contexts (Berkeley, Calif.)* 16, 2, https://doi.org/ 10.1177/1536504217714251 and [12] Rogers Brubaker. 2016. *Trans: Gender and Race in an Age of Unsettled Identities*. Gender and race in an age of unsettled identities Princeton University Press, Princeton.

[2] Think Grace Halsell’s *Soul Sister*, or John Howard Griffin’s *Black Like Me* for progressive attempts by whites to narrate the Black experience through “embodiment”; Ellen and William Craft’s life experience in using racial passing to escape enslavement; or minstrel shows and other forms of blackface as blunt racist acts.

[3] Theories by art historians about “reading images” were foundational to early computer vision and continue to be.

[4] Those interested in the genealogy of the pixel is invited to read Alvy Ray Smith’s beautiful book on the matter. “Pixels are invisible bits that represent visible waves. (….) There’s a common misconcenption that a pixel is a little square of color. But in fact, the pixel is a profound and abstract concept that binds our modern media world together.” P. 5, [25] Alvy Ray Smith. 2021. *A Biography of the Pixel* MIT Press. For the purposes of this paper, I will take a reductive definition of the pixel as the smallest resolved unit of a digital image. Greenland shows that pixels act hermeneutically because, “When combined in rasterized grids, pixels work together to create representations of reality that are comprehensible to the human eye.” P. 2, [26] Fiona A. Greenland. 2022. Pixel Politics and Satellite Interpretation in the Syrian War. *Media, Culture & Society* https://doi.org/ 10.1177/01634437221077169

phenomenon with difficult-to-differentiate units of meaning: No next-token prediction in the absence of semantic tokens.[5] All of this amplifies in digital photography, where the smallest unit of analysis, the pixel, contains little meaning to the human eye, but when considered in its spatial configuration, a lot of potential meaning for the computer. From this follow three assertions: A photograph is more proximate to an event than language; because of the way language and images work, language is "pre-labeled" in that it encodes mental models in ways that images do not, even though photographs also bear the marks of authorship and are culturally contingent; pictorial objects exist in the space of a linguistically disclosed world and yet function differently to language in ways that are meaningful for tackling social problems of sociological interest (such as: racial bias). Taken together, these three assertions reveal images as particularly complex sites of cultural meaning when compared to text. Add to this the history of racialized exploitation, stigmatization and surveillance through images and photography in particular [29-34], and we get a moral powder keg.

Returning to the lobster, this paper will think from a sociological standpoint through the question of morality and epistemology in images after the generative turn in AI. To do so, I draw up a three-part typology of epistemological problems afflicting generative AI:[6]

1. *Distance in a Representational Model of Data*: The main reason for using synthetic data is that it is understood to be a cheaper, convenient, privacy protective and more accessible stand-in for real world observations. This is based on a presumed proximity to the event, which in itself harbors both a promise and also a potential for violence.[7]
2. *Indexical Rupture*: The question of distance from the event can be further disambiguated. When the link to the real is severed, as in the case of deepfakes, what the philosopher Regina Rini calls the "epistemic backstop" to our testimonial practices becomes severed [36]. We can define this second problem as a problem of ruptured indexicality, where the link between what occurred and what is depicted terminates.[8]
3. *Performativity*: Philosophers of science and sociologists have studied how new forms of classification impact the populations so classified – Ian Hacking called this "looping effects" by which new types of humans are "made up" [8, 9, 38]. This problem links

[5] Those in the know may challenge me at this point by calling out the existence of visual transformers and token-based image representation. This is fair. See, for example: [28] Bichen Wu, Chenfeng Xu, Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Zhicheng Yan, Masayoshi Tomizuka, Joseph Gonzalez, Kurt Keutzer, and Peter Vajda. 2020. Visual Transformers: Token-based Image Representation and Processing for Computer Vision. In *Proceedings of the Computer Vision and Pattern Recognition*, 2020, Ithaca. Cornell University Library, https://arxiv.org/abs/2006.03677, [insert City of Publication],

[6] This typology is decidedly non-exhaustive and is also shaped by my disciplinary concerns as a sociologist.

[7] Notably absent from this piece is video surveillance used in modern warfare. There is a vast literature on the topic; I want to note here only a most recent piece: [35] Yuval Abraham 2024. *'Lavender': The AI machine directing Israel's bombing spree in Gaza*. City.

[8] On the question of indexicality and how it matters for analog versus digital images, see Gunning, who argues that the transition to the digital does not necessarily imply an indexical break. [37] Tom Gunning. 2004. What's the Point of an Index? or, Faking Photographs. *Nordicom review* 25, 1-2,

to, but is not identical to, the problem of recursion. Computer scientists have shown how a different kind of looping, that of recursion in training synthetic data on synthetic data, leads to irreversible defects in the resulting data [19, 39, 40].

After setting up the case and building on the significant literature in Science and Technology Studies on the politics of artifacts to show that photographs can be moral agents in the world, we will sketch out some representational and allocative harms [41] that emerge through these processes.

2 DO ARTIFACTS HAVE MORALS?

Photographs have been and continue to be crucial to projects of violence and the establishment of human difference [34, 42-46]. Their role in the history of policing and in the surveillance of deviant and immigrant populations is long [47-49]. Moreover, they are increasingly deployed as training data in automated systems [50, 51]. Photographs of faces as a subcategory of human portraits have a particularly concerning history that had its early hayday in the 19th century, including Alphonse Bertillon's standardized mugshots that redefined Parisian policing, the Italian Cesare Lombroso's portraits of criminal men, and the Brit Francis Galton's composite portraits of criminal types [52].

In his classic article, *Do Artifacts have Politics?,* Langdon Winner outlines two ways in which artifacts can have political qualities [53: 134]. The most obvious are instances in which the "intractable properties of certain kinds of technology are strongly (…) linked to particular institutionalized patterns of power and authority" [53: 134]. Take the example of the invention of mustard gas, the heinousness of which Benjamín Labatut described so effectively in his semi-fictionalized narrative of Fritz Haber, the Jewish German chemist whose wife, the chemist Clara Immerwahr, committed suicide to protest Haber's invention of the lethal gas [54]. He had continued to insist that technologies were morally neutral because they could be separated from their uses, inducing her self-murder. Haber would not live to see the use of a derivative of another one of his inventions in the mass murder of European Jews in the gas chambers: Cyclone B.

That a dangerous technology cannot be separated from its context is perhaps not coincidentally also a strong argument in AI Ethics, with researchers rightly pointing out the parallels between contemporary AI and historical forms of physiognomy, in which character traits are imputed based on facial features. This would make developments in artificial intelligence continuous with a physiognomic past, especially with regards to their shared essentializing and biological notions of ascriptive characteristic like race. Media studies scholars have in response argued for the banning of what they call "Physiognomic Artificial Intelligence" [55]. After all, in a context of doomsday rhetoric and exaggerated claims of rupture and novelty, it is decisive to also study continuities between analog and digital forms of data [56].

Stark and Hutson-style arguments are at the core of the European AI Act. The act, first proposed in April 2021, is unique in the world in terms of scope and goals. It "aims to ensure that AI systems placed on the European market and used in the EU are safe and respect fundamental rights and EU values."[9] In short, any company intending to bring a product deploying artificial intelligence to market in the EU must comply with the regulations of the act. The regulations are aimed at creating trustworthiness, protecting human rights, and mitigating risks. A core part of the regulatory framework is a risk-based assessment that defines and stratifies risk into four tiers: Unacceptable risk, High risk, Limited risk, and Minimal or no risk. An example for application that pose "unacceptable risk" are "AI systems considered a threat to people;" these "will be banned."[10] Such systems include "Biometric identification and categorization of people" as well as "real-time and remote biometric identification systems, such as facial recognition."[11]

Thus, the science and technology studies claim that artifacts have politics has been translated into regulation by policymakers who see images as a moral and ethical risk that requires intervention. While the space of AI Ethics is still very much in formation, groups of "moral entrepreneurs" are emerging. Such actors, whom we can define as wielding "their social influence to define ethical conduct" [2: 1], not only create and enforce norms in the larger public and political sphere, but perhaps more importantly are also crucial at defining what the exact "problem" is that has to be solved. Science and Technology Studies terminology would call this a problem of "translation": One in which an even objective and straightforward seeming problem – the decline of scallops in the St Brieuc bay [4]; the under-representation of people of color in image training sets – becomes framed in a certain way that excludes other ways of seeing.[12] Broadly, there have been four types of moral entrepreneurial responses to the problem of race in machine learning: Censorship and regulation, in which some data generation is forbidden; an idealist fairness approach that works through computational correctives; a more radical call for algorithmic reparations [57]; and a pragmatic approach of generating synthetic data to rebalance datasets. In the following, we will gesture at some actual examples of synthetic data generation and critically analyze them.[13]

[9] https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/

[10] https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[11] https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en

[12] Callon explains that we can reverse engineer the pipeline from public framing of a problem to its initial definition by thinking through the issue's problematization, interessement, enrolment and finally mobilization, where the particular framing and understanding of a technical issue is transmitted to the public [4] Michel Callon. 1984. Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The Sociological Review* 32, 1_suppl, https://doi.org/ 10.1111/j.1467-954X.1984.tb00113.x.

[13] Because we are sociologists and not computer scientists, and because the purpose of this piece is to carve out and theorize epistemological issues, the article only gestures at, but does not systematically study computer science publications in the field.

3 THE ROAD TO HELL IS PAVED WITH SYNTHETIC DATA

Winner-style arguments have much evolved since 1980, with the AI Ethics scholarship directly addressing the violent potential inherent in images. Scholars working on algorithmic justice have demonstrated that because many commercial machine learning datasets are not balanced for race and gender, they end up misclassifying non-white faces at high rates [58, 59]. In several well reported upon cases, this led innocent black men to be falsely arrested by the police [60].[14] One of the manifold issues in such cases of racial bias is the garbage in / garbage out problem [63] that could, ostensibly, be fixed by balancing the training sets by producing more "inclusive" data.

Given the paucity of available images of non-white faces on the internet, and the ethical issues involved in taking photos without people's permission, synthetic data appeared as a convenient possibility [13]. Let us return to the illustrations at the beginning of the piece: One of the examples in which researchers used generative modeling to solve problems of privacy, data integrity and a lack of sufficient data is the infamous racial domain translation paper. In a 2020 paper, Jiancheng Ge and coauthors decided that one might fix racial and ethnic distribution imbalances in machine learning training by generating more images of non-white persons [64]. Because they saw collecting new data as too costly, their proposal was to us a generative adversarial network (GAN). Deep generative models are a type of deep neural nets that interpret and then "model a probability distribution of the authentic training data" [65: 4]. In the process, random points are taken from the distribution and then mapped through a neural network. The goal is to get the thus generated distribution to match the actual underlying data distribution as closely as possible (ibid). Ge et al. applied this technology to their specific use case: changing white users to users of color, by "transfer(ring) the facial images of one race to corresponding images of other races, to facilitate the data augmentation to balance the ethnic distribution" (sic, ibid: 1).

The presumption underlying the process of multi-domain translation is that images can be translated between multiple and distinct domains. One of the papers debuting these skills was a 2017 International Conference on Computer Vision (ICCV) contribution on unpaired image-to-image translation that turned one "kind" of image (like an image of a horse) into another "kind" of image (like an image of a zebra, see figure 3).

[14] Black men are of course not the only targets – papers predicting "gay face" and "authoritarian face" have since emerged to great public fanfare: [61] Yilun Wang, Michal Kosinski, and Shinobu Kitayama. 2018. Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images. *Journal of personality and social psychology* 114, 2, https://doi.org/ 10.1037/pspa0000098, [62] S. H. R. Rasmussen, S. G. Ludeke, and R. Klemmensen. 2023. Using deep learning to predict ideology from facial photographs: expressions, beauty, and extra-facial information. *Sci Rep* 13, 1, (20230331), https://doi.org/ 10.1038/s41598-023-31796-1

Figure 3. Zhu et al, "Image Translation" [66], p. 2242

Put differently, the authors tried to learn mapping functions between two domains given a set of unpaired training samples [66].[15] Notably, while color and texture changes were successful, more sophisticated changes involving geometric transfers failed – figure 4 shows some degenerations, like the unsuccessful transfer from dog to cat, or from horse to zebra. Ge et al.'s work in this context is comparatively "successful" – it does not obviously "fail" in the same way as the horse --> zebra transfer fails, so let us investigate how they achieved what they refer to as "racial conversion."

Figure 12: Some failure cases of our method.

Figure 4. Zhu et al, "Image Translation" [66], p. 2249

[15] That the training samples are unpaired is crucial here – later models like CLIP were trained on paired text-image data, which erodes the fundamental distinction between labeled text (words, tokens) and unlabeled images (images).

The authors take as data source the "Racial Faces in-the-Wild" (sic) dataset, a dataset distilled from the MS-Celeb-1M dataset.[16] They then divided this dataset into four racial "domains" – "namely Caucasian, Asian, Indian and Africans" [64: 4]. A domain here refers to a set of images that are presumed to have the same characteristics. You will notice that the authors themselves tagged the photographs and dropped all ambiguous – read: possibly mixed-race observations – from the dataset. The authors also do not justify why they chose to divide their dataset into four racial categories, rather than fewer or more.[17] This leads to the multidimensional category of race collapsing into one reduced to visual phenotype, which has been rightfully criticized as an untenable form of operationalization [21, 68]. The authors then leverage the shared representations to "transform one racial domain images to other racial domains" [64: 1, sic] through their specific spin on data augmentation.

The research design itself, while not precisely analogous to blackface, is obviously offensive. But it is the specificity of synthetic data that makes this case study sociologically compelling: In the process of creating humans from existing representations of data, we are engaging in a Frankensteinian process of taking things apart and putting them together in combinations that go above and beyond earlier, pre-digital and pre-deep learning ways of doing composite typing, whether in criminal photography or for informant protection in ethnographic scholarship [69-71]. Given the long histories of racial classification being tied to life and death outcomes, and that battles over classification often included fears of racial "deception" that led to the policing of racial boundaries, such moves of statistical collection, fragmentation, and re-composition becomes especially odious when historically minoritized and persecuted bodies are concerned.

Now, what happens when one tries to provide racially balanced representation but also does so in ways that renounce a representational model of data? In short, what are the consequences of this double crisis of representation?

4 RUPTURED INDEXICALITY: THE DOUBLE CRISIS OF REPRESENTATION

The strength and weakness of synthetic data is that it is no longer tethered to a representational model of data. If we define non-synthetic data as "a record created through observation or measurement," then "the datum represents its object, points to it, and substitutes it in subsequent operations. In other words, the datum serves as a sign" [15: 2]. This presumes a direct and traceable relationship between the object in the world, and our

[16] https://exposing.ai/msceleb/

[17] This, the lack of explanation of one's operationalization, is not uncommon – building on the work of Zaid Khan and Yun Fu, I find in a spin-off project that studies how computer vision engineers define and delimit racial categories that these are ill-defined, unstable temporally and geographically and have a problematic history of scientific use. See [20] Zaid Khan, and Yun Fu 2021. *One Label, One Billion Faces: Usage and Consistency of Racial Categories in Computer Vision*. City, [67] A. K. M. Skarpelis. 2024. Beyond Physiognomic AI: The Long 20th Century of Racial Relativism. In short, computer vision engineers have little idea what race is and make badly reasoned *ad hoc* decisions.

recording of it.[18] Synthetic data in its multiple forms[19] dissolves this relationship: It preserves statistical properties and internal correlations while lacking representational correspondence that would connect the datum to the real world. Whereas previously, a "good" image was one that represented reality faithfully, the severing of the representational link means that a "good" image now is one that meets the requirements of the intended task. Offenhuber contends that an image's utility thus becomes tied to a relational and no longer a representational logic. Ge et al's model delivers on that account – it yields "good" images, because a recognition model trained on these new and augmented humans improves the recognition rates of African-Americans [64, 2].

Whether the images themselves are authentic, i.e., point at people in the real world, matters little to the model and to the use cases. This kind of racial transformation seems like a marginal problem, but what should we make of advertising and modeling agencies that use this technology to "diversify" their array of models by arbitrarily combining body parts, especially facial features, and skin tones? Seeing that the modeling industry has historically had a strong casting preference for white models, is this the racialized equivalent of green- or pink-washing, in which industries pay lip service and act deceptively to convince their customers that the company holds the moral highground?

Images used as input to train models are, fundamentally, "operational" images. Operational images are images that have a purpose, that are "not made for humans but for machines, used as the basis for various computational operations (Parikka 2023)" [15: 6]. Photographs of bombs dropped in the 1991 Gulf War are but one example that Harun Farocki, the originator of the phrase, used in his *Eye/Machine*[20] video installation. In their purposeful and cold use, these images often "abandon(s) the pictorial and representational function of images" [15: 6]. Reinterpreting Alphonse Bertillon's mugshot work, the anthropologist Zeynep Gürsel contends that the great innovation of his work was not the standardization of photographic capture, but instead was the taming of volume through organization [73]. The photographs themselves were not extraordinary; the retrieval system that they became a part of was the novelty that allowed the chief of criminal identification to "take the bodies out of the archive" (ibid). The photos by that point were almost an afterthought.

That we might not be able to make sense of the output of a specific convolutional layer in a neural net is therefore not a problem, representational logic be damned – as long as the relational logic holds, we are in business. Operational images of course are not new – paper maps used in warfare have always been operational and also two dimensional (or even three dimensional). Similarly, the production of these images has been remarkably consistent, in

[18] I have not yet engaged with Weatherby and Justie's "Indexical AI," and will do that in the next iteration of the paper. See: [72] Leif Weatherby, and Brian Justie. 2022. Indexical AI. *Critical inquiry* 48, 2, https://doi.org/ 10.1086/717312
[19] Offenhuber provides seven types in his typology: algorithmically generated data; imputed data; harmonized data; augmented and hybridized data; mimicry data; data inpainting; and synthetic training data. [15] Dietmar Offenhuber. 2024. Shapes and Frictions of Synthetic Data. *Big Data & Society* 11, 2, https://doi.org/ 10.1177/20539517241249390
[20] https://www.harunfarocki.de/installations/2000s/2000/eye-machine.html

that photographic studios have been standardized: Take Bertillon in Paris; take the Nazis' Central Immigration Office's training manual for field photographers that shows how to set up a mobile photographic studio (see figures 5-6).

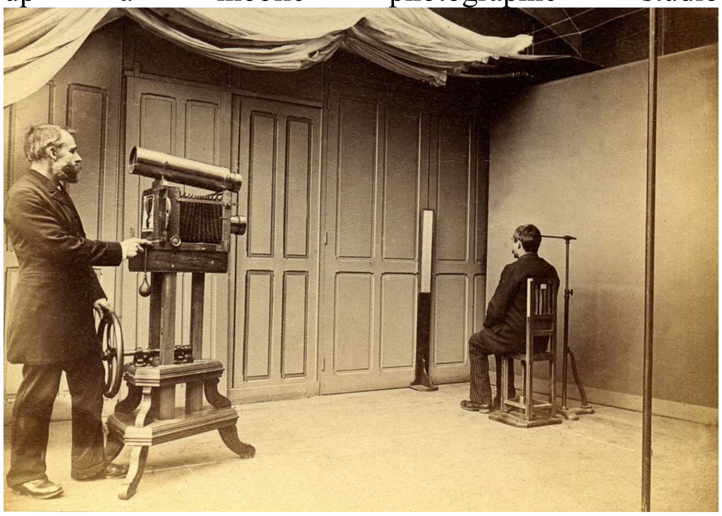

Figure 5. Bertillon's System for Photographing Criminals

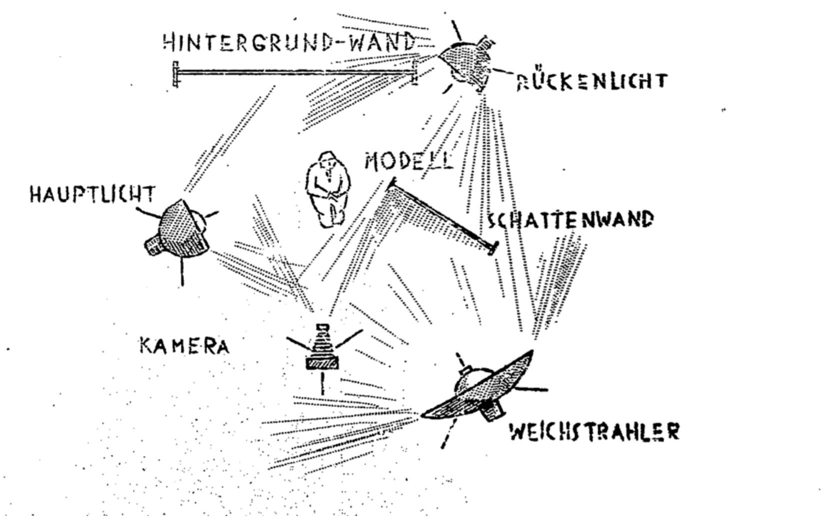

Figure 6. Schutzstaffel photographic manual on how to set up a mobile photographic studio. Source: BArch NS2/161, p. 73.

Data, even when used for global projects, have complex attachments to place – they are always "assembled from heterogeneous sources, each with their own local conditions" [74: 1983]. Consider what this means when image-text-pairs are scraped from "the internet," compared to when a unique dataset is generated for a specific purpose. One instance of the latter are early face recognition image sets that often featured the computer science lab members, and therefore depicted almost exclusively male, and in the 1960s and 1970s, majority white faces. Loukissas' suggestion that we ought to engage with data *settings*, rather than study the *datasets* in isolation, is handy [75]. Critical AI scholars have done this to help us understand the values that are embedded in ostensibly objective types of data [76-82]. The hope here is that through a combination of transparency, sight, and proximity, the abject will be rendered visible, which will compel political action. Political theorists Bernardo Zacka and Jasmine English [83] refer to this as "the politics of sight" in their discussion of Timothy Pachirat's slaughterhouse film "Every Twelve Seconds."[21] Synthetic data, by removing the tie between physical object and its representation, forfeits this possibility: Generative AI destroys what the philosopher Regina Rini called an "epistemic backstop" [36].[22]

But testimony is not accessible even where we have an epistemic backstop. Todd Presner and team have shown for the case of Holocaust Memory that one danger of representing testimonial data is the reduction of these narratives to "bare data," in which human beings are reduced to data points [84].

5 DISCUSSION

What do we make of this? Are we dealing simply with a reproduction of whiteness and a global imposition of US values? Years ago, the sociologists Pierre Bourdieu and Loïc Wacquant pointed at a similar process of a "planetary doxa" through the worldwide spread of "American concerns and viewpoints" [85]. More recently, Jenka Gurfinkel [86] wrote about how generative AI reproduces American visual semantics – her visual example of the Spanish Conquistadors with perfectly white, American teeth, huddling together like a US teenage football team is striking. Does this mean that we are witnessing the "consistent reproduction of whiteness as a latent feature of dominant visual culture", what Offert and Phao called semantic compression [22, 87]? It strikes us that gaffes like the Black Nazi, the "sassy queer Black" AI assistant and the transracial data augmentation reveal processes that would seem to go beyond these perceptive takes on the "cunning of imperialist reason."

Certainly, values suffuse the data and thinking about garbage-in-garbage-out, annotations as cultural artifacts, and human in the loop processes should be the foundation of any critical AI analysis. At the same time, we need to push further. First, there is something meaningful in the differences of form and meaning between text and images, especially when it gets to multimodal models. These acts of translation need to be theorized beyond simply pointing at how meaning in images is generated by recourse to different types of text (metadata and labels), or what Yann LeCun called "cheating", by "using language as a crutch to help the

[21] We have come full circle to lobster torture.

[22] Gini writes: "Our reasonable trust in the testimony of others depends, to a surprising extent, on the regulative effects of the ever-present possibility of recordings of the events they testify about" (1).

deficiencies of our vision systems to learn good representations from images and video."[23] Second, we need to be more nuanced in our narratives of continuities and rupture. As sociologists, historians, and media theorists have argued, the continuities often are more relevant in explaining a phenomenon than are changes, but they are also harder to explain [37, 56, 88, 89]. Lastly, the question of performativity. Liv, the sassy queer Black AI assistant, a form of "digital blackface" created by Meta [1], is unlikely to result in the type of looping effects Ian Hacking describes – no group boundaries are changed, but they do produce very real effects on people that go beyond the representational harm of tacky stereotypes. Economic sociologists have articulated how new classification technologies reshape political life, for example by creating new affordances for political mobilization [8]. Such a move is still outstanding in the literature on AI ethics and representational harms, and we hope to have outline some ways in which these questions could be approached in the future.

REFERENCES

[1] Karen Attiah 2025. *I talked to Meta's Black AI character. Here's what she told me. Is this the new era of digital blackface?*

[2] Mira D. Vale. 2024. Moral Entrepreneurship and the Ethics of Artificial Intelligence in Digital Psychiatry. *Socius* 10https://doi.org/ 10.1177/23780231241259641

[3] Luke Stark 2024. *Animation and Artificial Intelligence*. ACM, Rio de Janeiro.

[4] Michel Callon. 1984. Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The Sociological Review* 32, https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

[5] Bruno Latour, Steve Woolgar, and Jonas Salk. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, Princeton, N.J.

[6] Michael Lynch. 1988. The Externalized Retina: Selection and Mathematization in the Visual Documentation of Objects in the Life Sciences. *Human studies* 11, 2/3, https://doi.org/10.1007/bf00177304

[7] Wendy Espeland, xa, Nelson, and Michael Sauder. 2007. Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology* 113, 1, https://doi.org/10.1086/517897

[8] Greta R. Krippner, and Daniel Hirschman. 2022. The Person of the Category: The Pricing of Risk and the Politics of Classification in Insurance and Credit. *Theory and Society* 51, 5, https://doi.org/10.1007/s11186-022-09500-5

[9] Barbara Kiviat. 2023. The Moral Affordances of Construing People as Cases: How Algorithms and the Data They Depend on Obscure Narrative and Noncomparative Justice. *Sociological Theory* 41, 3, https://doi.org/10.1177/07352751231186797

[10] Jiancheng Ge, Weihong Deng, Mei Wang, and Jiani Hu 2020. *FGAN: Fan-Shaped GAN for Racial Transformation*. IEEE

[11] Mansi Hitesh. 2024. Troubling Transracialism: A Transnational Perspective. *Feminist theory* https://doi.org/10.1177/14647001241231717

[23] Source: https://lexfridman.com/yann-lecun-3-transcript/

[12] Ann Morning. 2017. Race and Rachel Doležal. *Contexts (Berkeley, Calif.)* 16, 2, https://doi.org/10.1177/1536504217714251

[13] Rogers Brubaker. 2016. *Trans: Gender and Race in an Age of Unsettled Identities*. Gender and race in an age of unsettled identities Princeton University Press, Princeton.

[14] Benjamin N. Jacobsen. 2023. Machine Learning and the Politics of Synthetic Data. *Big Data & Society* 10, 1, https://doi.org/10.1177/20539517221145372

[15] Benjamin N. Jacobsen. 2024. The Logic of the Synthetic Supplement in Algorithmic Societies. *Theory, Culture & Society* https://doi.org/10.1177/02632764231225768

[16] Dietmar Offenhuber. 2024. Shapes and Frictions of Synthetic Data. *Big Data & Society* 11, 2, https://doi.org/10.1177/20539517241249390

[17] Paula Helm, Benjamin Lipp, and Roser Pujadas. 2024. Generating Reality and Silencing Debate: Synthetic Data as Discursive Device. *Big Data & Society* 11, 2, https://doi.org/10.1177/20539517241249447

[18] Nima Kordzadeh, and Maryam Ghasemaghaei. 2022. Algorithmic Bias: Review, Synthesis, and Future Research Directions. *European Journal of Information Systems* 31, 3, https://doi.org/10.1080/0960085X.2021.1927212

[19] Cedric Deslandes Whitney, and Justin Norman. 2024. Real Risks of Fake Data: Synthetic Data, Diversity-Washing and Consent Circumvention. In *Proceedings of the Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2024, Rio de Janeiro, Brazil. Association for Computing Machinery, Rio De Janeiro, 1733–1744. https://doi.org/10.1145/3630106.3659002

[20] Sierra Wyllie, Ilia Shumailov, and Nicolas Papernot. 2024. Fairness Feedback Loops: Training on Synthetic Data Amplifies Bias. In *Proceedings of the Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2024, Rio de Janeiro, Brazil. Association for Computing Machinery, [insert City of Publication], 2113–2147. https://doi.org/10.1145/3630106.3659029

[21] Zaid Khan, and Yun Fu 2021. *One Label, One Billion Faces: Usage and Consistency of Racial Categories in Computer Vision*.

[22] Sebastian Benthall, and Bruce D. Haynes. 2019. Racial Categories in Machine Learning. In *Proceedings of the Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, Atlanta, GA, USA. Association for Computing Machinery, [insert City of Publication], 289-298. https://doi.org/10.1145/3287560.3287575

[23] Fabian Offert, and THao Phan. 2024. A Sign That Spells. *Journal of Digital Social Research* 6, 4, https://doi.org/ 10.33621/jdsr.v6i440462

[24] David Foster Wallace 2004. *Consider the Lobster*. Gourmet Magazine.

[25] M. F. K. Fisher 1941. *Consider the Oyster*. Gourmet Magazine.

[26] Alvy Ray Smith. 2021. *A Biography of the Pixel* MIT Press.

[27] Fiona A. Greenland. 2022. Pixel Politics and Satellite Interpretation in the Syrian War. *Media, Culture & Society* https://doi.org/10.1177/01634437221077169

[28] H. Talbot. 1844. The Pencil of Nature.

[29] Bichen Wu, Chenfeng Xu, Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Zhicheng Yan, Masayoshi Tomizuka, Joseph Gonzalez, Kurt Keutzer, and Peter Vajda. 2020. Visual Transformers: Token-based Image Representation and Processing for Computer Vision. In *Proceedings of the Computer Vision and Pattern Recognition*, 2020, Ithaca. Cornell University Library, https://arxiv.org/abs/2006.03677

[30] Sherrin Berezowsky. 2015. Statistical Criticism and the Eminent Man in Francis Galton's Hereditary Genius. *Victorian Literature and Culture* 43, 4, https://doi.org/10.1017/s1060150315000273

[31] H. Rose, and S. Rose. 2011. The Legacies of Francis Galton. *Lancet* 377, 9775, https://doi.org/10.1016/s0140-6736(11)60560-6

[32] Simone Browne. 2010. Digital Epidermalization: Race, Identity and Biometrics. *Critical sociology* 36, 1, https://doi.org/10.1177/0896920509347144

[33] Simone Browne. 2015. *Dark Matters: On the Surveillance of Blackness*. On the surveillance of blackness Duke University Press, Durham.

[34] Tina Campt. 2017. *Listening to Images*. Duke University Press, Durham.

[35] Ariella Azoulay. 2019. *Potential History: Unlearning Imperialism*. Verso, London

[36] Yuval Abraham 2024. *'Lavender': The AI machine directing Israel's bombing spree in Gaza*. 972 Magazine.

[37] Regina Rini. 2020. Deepfakes and the Epistemic Backstop. *Philosophers' imprint* 20

[38] Tom Gunning. 2004. What's the Point of an Index? or, Faking Photographs. *Nordicom review* 25

[39] Ian Hacking. 2006. Making up People. *London Review of Books* 28

[40] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. 2024. AI Models Collapse when Trained on Recursively Generated Data. *Nature* 631, 8022, https://doi.org/10.1038/s41586-024-07566-y

[41] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Gal Yarin, Nicolas Papernot, and Ross Anderson 2024. *The Curse of Recursion: Training on Generated Data Makes Models Forget*. Cornell University Library, arXiv.org,

[42] Jared Katzman, Angelina Wang, Morgan Scheuerman, Blodgett Su Lin, Kristen Laird, Hanna Wallach, and Solon Barocas 2023. *Taxonomizing and Measuring Representational Harms: A Look at Image Tagging*. Cornell University Library, arXiv.org

[43] A. K. M. Skarpelis. 2023. Horror Vacui: Racial Misalignment, Symbolic Repair, and Imperial Legitimation in German National Socialist Portrait Photography. *The American Journal of Sociology* 129, 2, https://doi.org/10.1086/727562

[44] Kevin P. Coleman, and Daniel James. 2021. Capitalism and the Camera: Essays on Photography and Extraction. Essays on photography and extraction Verso, London.

[45] John Tagg. 1993. The Burden of Representation: Essays on Photographies and Histories. University of Minnesota Press, Minneapolis.

[46] Lorena Rizzo. 2013. Shades of Empire: Police Photography in German South-West Africa. Visual anthropology (Journal) 26, 4, https://doi.org/10.1080/08949468.2013.804701

[47] Zeynep Devrim Gürsel. 2021. Bertillon, Ravachol and the Explosive Potential of Police Portraiture. History of Photography 45, 3-4, https://doi.org/10.1080/03087298.2021.2145708

[48] Lily Cho. 2021. Mass Capture: Chinese Head Tax and the Making of Non-Citizens. McGill-Queen's University Press, Montreal.

[49] Antigoni Memou. 2020. Photography, Borders and the Fear of the 'Racialised Other'. Liminalities 16, 2,

[50] Petra Molnar, and E. Tendayi Achiume. 2024. The Walls Have Eyes: Surviving Migration in the Age of Artificial Intelligence. The New Press, New York.

[51] Kelly A. Gates. 2011. Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance. NYU Press, New York.

[52] Monika Zalnieriute. 2021. Burning Bridges: The Automated Facial Recognition Technology and Public Space Surveillance in the Modern State. The Columbia science and technology law review 22, 2, https://doi.org/10.52214/stlr.v22i2.8666

[53] Leszek Brogowski. 2013. La photographie composite de Francis Galton, son protocole et son flou. L'épistémologie d'un protocole. In Proceedings of the Protocole & photographie contemporaine : colloque, Saint-Étienne, 8 - 10 novembre 2012, organisé par le Centre interdisciplinaire d'études et de recherches sur l'expression contemporaine; avec la collaboration scientifique du Lieven Gevaert research centre for photography; sous la direction de Danièle Méaux, Publications de l'Université de Saint-Étienne, 2013, Saint-Étienne, [insert City of Publication], 219-237.

[54] Langdon Winner. 1980. Do Artifacts Have Politics? Daedalus 109, 1,

[55] Benjamín Labatut. 2020. When We Cease to Understand the World. New York Review Books, New York.

[56] Luke Stark, and Jevan Hutson. 2021. Physiognomic Artificial Intelligence. Fordham Intell. Prop. Media & Ent. LJ 32

[57] Angele Christin. 2020. What Data Can Do: A Typology of Mechanisms. International journal of communication (Online)

[58] Jenny L. Davis, Apryl Williams, and Michael W. Yang. 2021. Algorithmic Reparation. Big Data & Society 8, 2, https://doi.org/10.1177/20539517211044808

[59] Joy Buolamwini, and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research: Conference on Fairness, Accountability, and Transparency 81

[60] Safiya Umoja Noble. 2018. Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press, New York.

[61] Kashmir Hill 2020. Wrongfully Accused by an Algorithm. City.

[62] Yilun Wang, Michal Kosinski, and Shinobu Kitayama. 2018. Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images. Journal of personality and social psychology 114, 2, https://doi.org/10.1037/pspa0000098

[63] S. H. R. Rasmussen, S. G. Ludeke, and R. Klemmensen. 2023. Using deep learning to predict ideology from facial photographs: expressions, beauty, and extra-facial information. Sci Rep 13, 1, (20230331), https://doi.org/10.1038/s41598-023-31796-1

[64] R. Stuart Geiger, Dominique Cope, Jamie Ip, Marsha Lotosh, Aayush Shah, Jenny Weng, and Rebekah Tang. 2021. "Garbage In, Garbage Out" Revisited: What do Machine Learning Application Papers Report About Human-Labeled Training Data? Quantitative Science Studies 2, 3, https://doi.org/10.1162/qss_a_00144

[65] Fadi Boutros, Vitomir Struc, Julian Fierrez, and Naser Damer. 2023. Synthetic Data for Face Recognition: Current State and Future Prospects. Image and vision computing 135 https://doi.org/10.1016/j.imavis.2023.104688

[66] Zhu Jun-Yan, Park Taesung, Phillip Isola, and Alexei A. Efros 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. IEEE, City.

[67] A. K. M. Skarpelis. 2024. Beyond Physiognomic AI: The Long 20th Century of Racial Relativism.

[68] Wendy D. Roth. 2016. The Multiple Dimensions of Race. Ethnic and Racial Studies 39, 8, https://doi.org/10.1080/01419870.2016.1140793

[69] Theodore M. Porter. 2020. The Rise of Statistical Thinking, 1820–1900. (New edition. ed.). Princeton University Press, Princeton.

[70] Noah Amir Arjomand. 2022. Empirical Fiction: Composite Character Narratives in Analytical Sociology. The American Sociologist https://doi.org/10.1007/s12108-022-09546-z

[71] Steven Ruggles, Catherine Fitch, Diana Magnuson, and Jonathan Schroeder. 2019. Differential Privacy and Census Data: Implications for Social and Economic Research. AEA papers and proceedings 109

https://doi.org/10.1257/pandp.20191107

[72] Leif Weatherby, and Brian Justie. 2022. Indexical AI. Critical inquiry 48, 2, https://doi.org/10.1086/717312

[73] Zeynep Devrim Gürsel. 2024. Photography as Border-Making. In Proceedings of the Leonard A. Lauder Research Center Talks, 2024, The Metropolitan Museum of Art, [insert City of Publication],

[74] Claire Tupling. 2021. All data are local: thinking critically in a data-driven society. Information, Communication & Society 24, 13, https://doi.org/10.1080/1369118X.2020.1817523

[75] Yanni Alexander Loukissas. 2019. All Data Are Local: Thinking Critically in a Data-Driven Society. The MIT Press (1. ed.). MIT Press, Cambridge.

[76] Chris Dulhanty, and Alexander Wong. 2019. Auditing ImageNet: Towards a Model-driven Framework for Annotating Demographic Attributes of Large-Scale Image Datasets. arXiv.org https://doi.org/10.48550/arxiv.1905.01347

[77] Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. 2021. On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet. Big Data & Society 8, 2, https://doi.org/10.1177/20539517211035955

[78] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. Proceedings of the ACM on Human-Computer Interaction 5, CSCW2, https://doi.org/10.1145/3476058

[79] Matthew Groh, Caleb Harris, Roxana Daneshjou, Omar Badri, and Arash Koochek. 2022. Towards Transparency in Dermatology Image Datasets with Skin Tone Annotations by Experts, Crowds, and an Algorithm.

[80] Abeba Birhane, Sepehr Dehdashtian, Vinay Prabhu, and Vishnu Boddeti 2024. The Dark Side of Dataset Scaling: Evaluating Racial Classification in Multimodal Models. City.

[81] Isak Engdahl. 2024. Agreements 'in the wild': Standards and alignment in machine learning benchmark dataset construction. Big Data & Society 11, 2, https://doi.org/10.1177/20539517241242457

[82] Kathleen C Fraser, and Svetlana Kiritchenko. 2024. Examining Gender and Racial Bias in Large Vision-Language Models Using a Novel Dataset of Parallel Images. arXiv preprint arXiv:2402.05779

[83] Jasmine English, and Bernardo Zacka. 2022. The Politics of Sight: Revisiting Timothy Pachirat's Every Twelve Seconds. American Political Science Review 116, 3, https://doi.org/10.1017/S0003055421001192

[84] Todd Samuel Presner, and Anna Bonazzi. 2024. Ethics of the Algorithm: Digital Humanities and Holocaust Memory. Digital humanities and Holocaust memory Princeton University Press, Princeton, New Jersey.

[85] Pierre Bourdieu, and Loïc Wacquant. 1999. On the Cunning of Imperialist Reason. Theory, Culture & Society 16, 1, https://doi.org/10.1177/026327699016001003

[86] Jenka Gurfinkel 2023. AI and the American Smile: How AI misrepresents culture through a facial expression. City.

[87] Thao Phan, and Scott Wark. 2021. Racial formations as data formations. Big Data & Society 8, 2, https://doi.org/10.1177/20539517211046377

[88] Orlando Patterson. 2004. Culture and Continuity: Causal Structures in Socio-Cultural Persistence Matters of Culture: Cultural Sociology in Practice, Cambridge University Press, Cambridge, UK.

[89] Caroline Walker Bynum. 1998. Metamorphosis, or Gerald and the Werewolf. Speculum 73, 4, https://doi.org/10.2307/2887366