

Fabian Offert

On the Concept of History (in Foundation Models)

2023

<https://doi.org/10.25969/mediarep/22316>

Veröffentlichungsversion / published version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Offert, Fabian: On the Concept of History (in Foundation Models). In: *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft*. Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, Jg. 19 (2023), Nr. 1, S. 121–134. DOI: <https://doi.org/10.25969/mediarep/22316>.

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under a Deposit License (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual, and limited right for using this document. This document is solely intended for your personal, non-commercial use. All copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute, or otherwise use the document in public.

By using this particular document, you accept the conditions of use stated above.

Fabian Offert

On the Concept of History (in Foundation Models)

Abstract: What is the concept of history inherent in contemporary models of visual culture like CLIP and DALL·E 2? This essay argues that, counter to the corporate interests behind such models, any understanding of history facilitated by them must be heavily politicized. This, the essay contends, is a result of a significant technical dependency on traditional forms of (re-)mediation. Polemically, for CLIP and CLIP-dependent generative models, the recent past is literally black and white, and the distant past is actually made of marble. Moreover, proprietary models like DALL·E 2 are intentionally cut off from the historical record in multiple ways as they are supposed to remain politically neutral and culturally agnostic. One of the many consequences is a (visual) world in which, for instance, fascism can never return because it is, paradoxically at the same time, censored (we cannot talk about it), remediated (it is safely confined to a black-and-white media prison), and erased (from the historical record).

Introduction

Any sufficiently complex technical object that exists in time has, in a sense, a concept of history: a way that the past continues to exist for it, with contingencies and omissions specific to its place and role in the world. This essay asks: what is the concept of history that emerges from a specific class of technical objects that have come to dominate the field of artificial intelligence, so-called ‘foundation models’? Do foundation models conceptualize the past?¹ What is the past for them? This question does not imply any intentionality, agency, or subjectivity on the part of the models under investigation. In fact, the argument that I would like to make is that a discernible concept of history does emerge from contemporary artificial intelligence systems despite an utter lack of intelligence

1 A technical rendering of the same question is: Do transformers learn world models or surface statistics (cf. LI 2023) of historical time?

in the general sense. The question, in other words, is entirely non-philosophical and non-speculative. It is exactly not ‘what is it like to be’ a foundation model. Instead, it could be rephrased as: as far as can be shown, is there internal consistency to the outputs of a foundation model when it is tasked with processing inputs related to the past? And if so, what are the structuring principles of these internally consistent outputs, and how do they relate to the structuring principles humans apply to the past to render it history?

My experimental close-readings of two such systems in particular, the CLIP model released by OpenAI in 2021 and the DALL-E 2 model released in 2022,² suggests that one of these structuring principles, and arguably the most significant at least for visual models, is a technically determined form of *remediation* (cf. BOLTER/GRUSIN 2000). Polemically, for CLIP and CLIP-dependent generative models, the recent past is literally black and white, and the distant past is actually made of marble. Given that CLIP, at the same time, *premediates* our future digital experience as a means of search, retrieval, and recommendation, this structuring principle of remediation then becomes ethically and politically relevant. As Alan Liu asks:

Today, the media question affects the sense of history to the core. [...] This is not just an abstract existential issue. It’s ethical, political, and in other ways critical, too. Have we chosen the best way to speak the sense of history today, and if so, for the benefit of whom? (LIU 2018: 2).

On the Concept of History

The ethical questions surrounding this ‘media question’ are maybe nowhere as obvious as in the digitization of the testimonies of those who survived the Holocaust (cf. WALDEN/MARRISON 2023). Projects like *Dimensions in Testimony*, which is funded by the USC Shoah Foundation, have started to go beyond the mere recording of testimonies, attempting to emulate their performative quality, the significant experience of sharing a moment in space and time, with the help of artificial intelligence. As the project website states:

Dimensions in Testimony enables people to ask questions that prompt real-time responses from pre-recorded video interviews with Holocaust survivors and other witnesses to genocide. The pioneering project integrates advanced filming techniques, specialized display technologies and next generation natural language processing to create an interactive biography (USC SHOAH FOUNDATION 2023: n.pag.).

2 Although CLIP and DALL-E 2 have been released separately, DALL-E 2 heavily depends on CLIP embeddings which guide the training process. See RAMESH et al. 2022 for details.

Todd Presner (2022) has pointed out the dilemma that such projects find themselves in. Humans, he argues, “are no longer (centrally) part of the creation of digital cultural memory”. Instead, through established and AI-enhanced technologies of montage, individual testimonies, once irreversibly tied to an individual human life, become disembodied. If the duty to keep these testimonies accessible for future generations warrants these technological interventions – “that Auschwitz not happens again”,³ in Adorno’s words – is an open question. Irrespective of such ethical considerations, projects like *Dimensions in Testimony* point to a fundamental media-theoretical question about the concept of history: What is the imprint that a specific technology leaves on history? More precisely, what, if anything, does artificial intelligence ‘add’ to an already (re-)mediated past?⁴

Here, we need to turn to Walter Benjamin’s text *Über den Begriff der Geschichte* (1974a) that the title of this essay takes inspiration from. Years of scholarly debate on Benjamin’s writings⁵ have made it unnecessary to introduce its premise here, or comment on the unusual synthesis of materialist and theological thought that it embodies. Instead, I would like to point out an almost trivial similarity between *Über den Begriff der Geschichte* and Benjamin’s other widely read essay on the *Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit* (1974b). Famously, in *Über den Begriff der Geschichte*, Benjamin writes: “To articulate the past historically does not mean to recognize it ‘the way it really was [...]’. It means to seize hold of a memory as it flashes up at a moment of danger”.⁶ Previously, in the *Kunstwerk*-essay, Benjamin had argued that the political potential of film derives from its power to produce abrupt cuts, and thus ‘chocks’ the viewer into a different mode of thinking. In other words, for Benjamin, the condition under which history becomes possible, the “moment of danger”, is the condition that film emulates. In both cases, awareness and insight depend on a moment of immediacy, and in both cases this moment of immediacy must be actively captured and repurposed for a progressive (Marxist) agenda before it falls into the hands of the fascists. There is thus, for Benjamin, a structural similarity between history as a memory that “flashes up”, that emerges from, and is actualized by, a moment of crisis, and the specific ways in which technology mediates our experience of the present

3 Translation and paraphrase by the author; original: “Die Forderung, daß Auschwitz nicht noch einmal sei, ist die allererste an Erziehung”, ADORNO 1970: 135.

4 Trivially, the past can only ever ‘reach’ us in mediated form. In the context of foundation models, for all but the most recent past, this also implies remediation, as foundation models only operate on digital (i.e., digitized or born-digital) data. While such earlier ‘layers’ of remediation have interesting media-theoretical implications of their own (cf. SEREXHE 2013) they are irrelevant in the context of this essay, which is concerned with the ‘surplus’ remediation introduced by foundation models exclusively.

5 See LÖWY 2005 for a good overview.

6 Translation by the author; original: “Vergangenes historisch zu artikulieren heißt nicht, es zu erkennen, ‘wie es denn eigentlich gewesen ist [...]’. Es heißt, sich einer Erinnerung bemächtigen, wie sie im Augenblick einer Gefahr aufblitzt”, BENJAMIN 1974a: 695.

world, and thus shapes our political views of it. Crucially, history and technology manifest themselves as a specific way of seeing.

What I am suggesting here, then, is not that we should ‘apply’ Benjamin’s concept of history to artificial intelligence systems. On the contrary: One of the reasons why the field of ‘critical AI studies’ has not had the impact that one would expect given the oversized importance of artificial intelligence research in computer science, is its insistence on resorting to traditional humanist theoretical frameworks and concepts that simply do not suffice anymore. Instead, I would like to propose, exactly with Benjamin, that we have to carve out the extremely specific, borderline idiosyncratic ways of seeing that artificial intelligence systems bring to the table where they are tasked with processing, or producing, an already mediated past. Again, more precisely: As the past is remediated through contemporary artificial intelligence systems, is the concept of history that emerges from this process of remediation different from the concept of history that emerges from the always already (re-)mediated data on its own? What, in other words, is the ‘surplus remediation’ inherent a foundation model’s specific way of seeing?

CLIP vs. DALL·E 2

Foundation model is a term introduced by a collective of researchers at the Stanford HAI institute in 2021 (cf. BOMMASANI et al. 2021). It basically means models that are a) very large, and b) that can be used for a variety of ‘downstream’ tasks. The vision model CLIP (contrastive language-image pre-training, cf. RADFORD et al. 2021), first released in 2021 by OpenAI, is such a foundation model. Outside the technical community, its innovations were somewhat obscured by the concurrent release of the DALL·E model, and later overshadowed by DALL·E’s successor, DALL·E 2 (cf. RAMESH et al. 2022) and the language model GPT-3.

CLIP – other than both iterations of DALL·E, as well as GPT-3 – is not a generative model. It does not produce images or text, but it connects them. More precisely, CLIP learns from images in context by projecting an image and its context into a common ‘embedding space’. The ‘context’ here could be an image caption, a so-called ‘alt text’ which describes the image in case it is not loaded properly and to accommodate people with screen readers, or simply a news article that the image illustrates. A fully trained CLIP model, then, consists of a high-dimensional vector space, or embedding space, in which words and images that are related can be found close together. Similarity between image and text is thus modeled as spatial proximity (this is true for all embedding models, be it just words, just images, or both, such as in the case of CLIP). While CLIP was

originally designed for zero-shot image labeling,⁷ it also facilitates what computer scientists call ‘image retrieval’ (this exemplifies its ‘foundation’ character): finding specific images within an unlabeled corpus of images based on visual or textual prompts. The user can provide CLIP with an image and it will look for similar images, or they can provide it with a prompt and it will look for images corresponding to this prompt – in any corpus of images. Given that the training corpus for CLIP is largely unknown,⁸ it seems futile to attempt to construct a somewhat empirical basis for our claims. And yet, there are two ways to study CLIP’s concept of history empirically.

Attribution by Proxy

The first way we could call ‘attribution by proxy’. While we do not know what CLIP was trained on, we can still ‘ask’ it for things *in terms* of specific collections of images. It is exactly this aspect of CLIP – the universality of its embeddings – that makes it so powerful as a retrieval engine. The following examples were tested with a custom CLIP-based search engine called *imgs.ai* (cf. OFFERT/BELL 2023), which indexes museum collections in the public domain.

Diego Velázquez’ 1656 painting *Las Meninas* is one of the most discussed pictures of art history. Using approaches from computer vision preceding CLIP, what can we say about this picture? We might be able to determine the number of people in the picture with the help of a pre-trained and/or fine-tuned face detection network. We might confirm the existence of certain image objects – an easel, a dog, or other paintings – with the help of an object detection network. We might even be able to estimate the gaze direction of some of the characters in the picture. But under no circumstances could we infer the play on representation that the picture embodies, the fact that it is, with William J.T. Mitchell, a “metapicture” (MITCHELL 1995: 35), a picture about pictures, a representation of (the concept of) representation.

In contrast, if we run an *imgs.ai* search for “Las Meninas” on the collection of the Museum of Modern Art, New York, an institution that does *not* have the famous painting in its collection (which is kept in the Prado in Madrid), the results are surprisingly ‘accurate’ and show the conceptual depth that CLIP allows the user to access. Among them are two photographic works, Joel

7 The technical term ‘zero-shot image labeling’ refers to the captioning of images without further training or fine-tuning a model on the dataset that contains them.

8 Here, I am referring to the specific, proprietary pre-trained model released by OpenAI in 2021. Since then, there have been multiple attempts to replicate CLIP in an open-source context. See, for instance, the OPENCLIP approach proposed by CHERTI et al. 2022, and research done at LAION to produce efficient pre-trained OPENCLIP models: <https://laion.ai/blog/large-openclip/> [accessed February 16, 2023].

Meyerowitz’ “Untitled” from “The French Portfolio” (1980, Fig. 1) and Robert Doisneau’s *La Dame Indignée* (1948, fig. 2). Both are explicit plays on representation, and both clearly pick up on the same themes as “Las Meninas”, especially the question of the *gaze* relation between people in, and people before the image, to use George Didi-Huberman’s (2004) term.



Figure 1: Joel Meyerowitz, „Untitled” from “The French Portfolio” (1980). Museum of Modern Art, New York



Figure 2: Robert Doisneau, *La Dame Indignée* (1948). Museum of Modern Art, New York

Replacing art history with history proper, and also going back to the ethical and political stakes of automated vision, we can query this same collection for “images of the Holocaust”. And the results tell us that, yes, CLIP ‘knows’ – too

well – what we are talking about. On the one hand, the model will suggest those few images in the MoMA collection that are historically linked to the query, for instance photographs by the U.S. Army Signal Corps which played an important role in documenting the atrocities of the Germans. But on the other hand, it will exemplify a much more abstract knowledge about visual Holocaust memory. Suggested results include a photograph by Bruce Davidson, shot on the set of the war film *Lost Command in Spain* in the 1960s (fig. 3), a 1980 photograph by Aaron Siskind depicting volcanic lava (fig. 4), a collage made from stamps by Robert Watts in 1963 (fig. 5), or a 1995 photograph by Alexander Slussarev that shows several pairs of shoes (fig. 6). None of these pictures are historically related to the Holocaust, nor are they necessarily meant to evoke it, but all of them could be easily recontextualized with respect to the visual language of Holocaust cultural memory. Using the MoMA collection as a proxy, we can see how well CLIP has internalized this visual language. Moreover, far from just showing the unshowable, CLIP has clearly learned that this language operates metaphorically. But: the fact that all the results that CLIP proposes (not only those named above) are black-and-white photos already points to a significant limitation, a limitation that we can further explore by utilizing generative models.



Figure 3: Bruce Davidson, *Spain* (1965).
Museum of Modern Art, New York



Figure 4: Aaron Siskind, *Volcano 1* (1980).
Museum of Modern Art, New York



Figure 5: Robert Watts, *Yamflug / 5 Post 5* (1963). Museum of Modern Art, New York



Figure 6: Alexander Slussarev, *Untitled* (1995). Museum of Modern Art, New York

Generative Attribution

This second way of studying CLIP we could call ‘generative attribution’.⁹ It is made possible by the fact that CLIP, to a large part, determines the training of generative models like DALL·E and Stable Diffusion.

If you ask the generative model, DALL·E 2, for “a color photo of a fascist parade, 1935” it will not comply. “Fascism”, among many other political terms, was banned by OpenAI early on to mitigate the potential of their model – of which they were well aware – to produce politically, legally, or socially unacceptable material like deep fakes, pornography, or propaganda. Such safeguards are not in place in other models like Stable Diffusion but there exists a simple trick to circumvent DALL·E’s forced ‘neutrality’ as well. Intentionally misspelling “fascism” by leaving out the “s”¹⁰ will produce (a variation of) the image in figure 7:

9 The use of generative approaches to ‘open the black box’ of artificial intelligence has first been proposed in the field of explainable artificial intelligence. For an overview of its epistemic implications, cf. OFFERT/BELL 2021.

10 I have argued elsewhere (cf. OFFERT 2022) that this kind of ‘humanist hacking’ which resorts to metalanguage will become more common in the near future. In the meantime (early 2023), OpenAI has improved their safeguards and the ‘hack’ will not work anymore.

a vaguely Western European city with some sort of mass rally taking place, red flags raised, and ominous smoke emerging from a building in the background. DALL·E, in other words, despite its safeguards, ‘knows’ very well what 1935 fascism looks like – *to us*. The generated image has the appearance of a historical photograph not only for its subject but for its appearance; it shows the characteristic colors of early Kodachrome slide photography, with the red of the flags particularly standing out against an otherwise subdued sepia palette. This is how Nazi Germany appears in the photographs of Hugo Jäger, for instance, whose pre-war slide collection was acquired and popularized by *LIFE* magazine in the 1960s.¹¹¹



Figure 7: DALL·E generation for “a color photo of a fascist [sic] parade, 1935”, produced in October 2022. Note that this safeguard circumvention technique has been ‘fixed’ at the time of writing

What is remarkable about this generated image is not its accuracy in emulating a specific historical medium – this has been possible at least since the early days of style transfer ca. 2016 – but that it resorts to this specific historical medium by default. Nowhere in the prompt did we ask for early Kodachrome in particular. And it turns out that it is hard to get rid of, too. From experiments done on both DALL·E 2 and Stable Diffusion, it is difficult to impossible to produce color photographs of fascist parades, ca. 1935, that do *not* have the appearance of early Kodachrome, colorized black-and-white, or otherwise historically more or less accurate photographic techniques. Only through copious amounts of highly specific additional keywords or negative prompts is it possible to steer the model away from this particular aesthetic. There exists, in other words, a strong default in models like DALL·E that conjoins historical periods and historical media and

111 Jäger’s images are not reproduced in this essay for ethical reasons. See Cosgrove n.d. for a sample of his specific aesthetic facilitated by early Kodachrome film.

thus produces a (visual) world in which fascism can simply not return because it is safely confined to a black-and-white media prison.

Foundation Models as Contingency Machines

Of course, all of this is, in a way, not very surprising. The past, for us and the model, exists visually only through those historical media that we see emulated here. Media determine our situation, for better or worse, and it is hard for us, too, to picture the past alive. What we are asking for here are speculative images, visual evidence that does not align with the documents *or* monuments left to us. And yet, the current generation of foundation models can easily produce highly speculative images when the speculation is ‘semantic’, not ‘syntactic’. Contemporary generative models are famously able to generate entirely fictional images like the well-known “astronaut riding a horse on the moon”. While DALL·E 2, for instance, has no problem producing a cartoon image of a cat driving a car, a realistic color photograph of a cat driving a car – where the cat actually drives the car, paws on the steering wheel – again requires copious amounts of prompt engineering. In short: for visual foundation models, ‘semantic’ speculation is easy, ‘syntactic’ speculation is hard.

The flip side of this capability is that it cannot be switched off easily. In the case of proprietary models like DALL·E 2, which includes additional safeguards that are supposed to guarantee it remains ‘culturally agnostic’ (cf. CETINIC 2022), this has significant consequences. While ‘allowed’, *generally* historical prompts (including those originally hidden behind surface-level, i.e., prompt parsing safeguards, like “fascism”) are tied to specific forms of mediation, *specifically* historical prompts are decoupled from the event that they refer to and relegated to a world of fiction. Why? Because the model *must have an answer*. As for all foundation models, failure is not an option – there has to be a result, no matter how outrageous. Foundation models, in other words, are *contingency machines*.¹² DALL·E 2, in particular, fails to reproduce historical images without altering their meaning. The prompt “Laocoön and His Sons, between 27 BC and 68 AD”, which references the famous work central to European art history since Winckelmann, produces a serene image of a Black¹³ family with no trace of agony (fig. 8). The prompt “Tank Man, 1989”, which references the iconic photograph from the Chinese Tiananmen protests, produces an image of a soldier proudly looking at a tank (fig. 9), rather than a scene of radical civil disobedience.

12 There is an argument to be made here, too, that such models, following Barthes’ (1982) analysis of textual contingencies, produce an estranged machinic *realism*.

13 That the family is depicted as Black is a result of a superficial bias mitigation attempt by OpenAI that was exposed in 2022; see OFFERT/PHAN 2022 for details.



Figure 8: DALL-E generation for *Laocoön and his sons*, between 27 BC and 68 AD, produced in October 2022



Figure 9: DALL-E generation for *Tank Man*, 1989, produced in October 2022

Conclusion

Answering one of our initial questions – what, if anything, does artificial intelligence ‘add’ to an already mediated past? – we now have to state that artificial intelligence not only adds nothing, but it forecloses a political potential. Models like DALL-E 2 find themselves in a triple bind: they suffer from syntactic invariability in the case of *generally* historical prompts, semantic arbitrariness in the case of *specifically* historical prompts, and superficial, corporate censorship that affects both. The result is an implicitly politicized concept of history. In the most literal interpretation of the famous idea that history doesn’t repeat itself, the past can never be actualized and is eternally tied to a specific medium, while images that are already rendered into history are excluded from making an appearance by

simple corporate policy. Neither can history be made by actualizing the past for the present, nor can the already-historical past be summoned. One of the many consequences is a (visual) world in which fascism can simply not return because it is, paradoxically at the same time, censored (we cannot talk about it), remediated (it is safely confined to a black-and-white media prison), and erased (from the historical record). More generally, in embedding models, the fundamental principle of computation – that time must become space¹⁴ – is applied, wrongly, to historical time. Historical time, encoded as (embedding) space, has no gaps, and does not even allow for gaps. In embedding space, there are simply no dots left to connect.

Bibliography

- ADORNO, THEODOR W.: *Erziehung nach Auschwitz*. In: GERD KADELBACH (ed.): *Erziehung zur Mündigkeit: Vorträge und Gespräche mit Hellmuth Becker 1959-1969*. Frankfurt/M. [Suhrkamp] 1970, pp. 135-162
- BARTHES, ROLAND: *The Reality Effect*. In: TZVETAN TODOROV (ed.): *French Literary Theory Today: A Reader*. Cambridge [Cambridge University Press] 1982, pp. 11-17
- BENJAMIN, WALTER: *Über den Begriff der Geschichte*. In: *Gesammelte Schriften I.2*. Frankfurt /M. [Suhrkamp] 1974a, pp. 693-704
- BENJAMIN, WALTER: *Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit*. In: *Gesammelte Schriften I.2*. Frankfurt /M. [Suhrkamp] 1974b, pp. 471-508
- BOLTER, JAY D.; RICHARD GRUSIN: *Remediation: Understanding New Media*. Cambridge, MA [MIT Press] 2000
- BOMMASANI, RISHI; et al.: *On the Opportunities and Risks of Foundation Models*. *arXiv:2108.07258*. August 16, 2021. <https://arxiv.org/abs/2108.07258> [accessed February 16, 2023]
- CETINIC, EVA: *Multimodal Models as Cultural Snapshots*. Talk given at Ludwig Forum Aachen, November 18, 2022
- CHERTI, MEHDI; et al.: *Reproducible Scaling Laws for Contrastive Language-Image Learning*. *arXiv:2212.07143*. December 14, 2022. <https://arxiv.org/abs/2212.07143> [accessed February 16, 2023]
- COSGROVE, BEN: *A Brutal Pageantry: The Third Reich's Myth-Making Machinery, in Color*. In: *LIFE History*. No date. <https://www.life.com/history/a-brutal-pageantry-the-third-reichs-myth-making-machinery-in-color/> [accessed February 16, 2023]

14 As Sybille Krämer (2006: 99) summarizes Friedrich Kittler: “Wherever something is stored, a temporal process must be materialized as a spatial structure. Creating spatiality becomes the primary operation by which the two remaining functions of data processing – transporting and processing – become possible at all”.

- DIDI-HUBERMAN, GEORGES: *The Surviving Image: Phantoms of Time and Time of Phantoms: Aby Warburg's History of Art*. University Park [Pennsylvania State University Press] 2017
- USC SHOAH FOUNDATION: *Dimensions in Testimony*. <https://sfi.usc.edu/dit> [accessed February 16, 2023]
- KRÄMER, SYBILLE: The Cultural Techniques of Time Axis Manipulation: On Friedrich Kittler's Conception of Media. In: *Theory, Culture & Society*, 23(7-8), 2006, pp. 93-109
- LI, KENNETH: Do Large Language Models Learn World Models or Just Surface Statistics? In: *The Gradient*, 2023. <https://thegradient.pub/othello/> [accessed February 16, 2023]
- LIU, ALAN: *Friending the Past: The Sense of History in the Digital Age*. Chicago [University of Chicago Press] 2018
- MITCHELL, WILLIAM J.T.: *Picture Theory: Essays on Verbal and Visual Representation*. Chicago [University of Chicago Press] 1995
- LÖWY, MICHAEL: *Fire Alarm: Reading Walter Benjamin's 'On the Concept of History'*. London [Verso] 2005
- OFFERT, FABIAN: On the Emergence of General Computation from Artificial Intelligence. In: *Zentralwerkstatt*. December 5, 2023. <https://zentralwerkstatt.org/blog/on-the-emergence-of-general-computation-from-artificial-intelligence> [accessed February 16, 2023]
- OFFERT, FABIAN; PETER BELL: Perceptual Bias and Technical Metapictures: Critical Machine Vision as a Humanities Challenge. In: *AI & Society*, 36, 2021, pp. 1133-1144
- OFFERT, FABIAN; PETER BELL: mgs.ai: A Deep Visual Search Engine for Digital Art History. In: *International Journal for Digital Art History*, 2023/forthcoming
- OFFERT, FABIAN; THAO PHAN: A Sign That Spells: DALL-E 2, Invisual Images and the Racial Politics of Feature Space. *arXiv:2211.06323*. October 26, 2022. <https://arxiv.org/abs/2211.06323> [accessed February 16, 2023]
- PRESNER, TODD: *Digitizing, Remediating, Remixing, and Reinterpreting Holocaust Memory*. Talk given at the University of California, Santa Barbara, May 10, 2022
- RADFORD, ALEC; et al.: Learning Transferable Visual Models from Natural Language Supervision. In: *International Conference on Machine Learning (ICML)*, 2021, pp. 8748-8763
- RAMESH, ADITYA; PRAFULLA DHARIWAL; ALEX NICHOL; CASEY CHU; MARK CHEN: Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv:2204.06125*. April 13, 2022. <https://arxiv.org/abs/2204.06125> [accessed February 16, 2023]
- SEREXHE, BERNHARD (ed.): *Konservierung Digitaler Kunst: Theorie und Praxis*. Vienna [Ambra V] 2013

WALDEN, VICTORIA GRACE; KATE MARRISON: *Recommendations for Digitally Recording, Recirculating, and Remixing Holocaust Testimony: Digital Holocaust Memory Project Report.* Sussex [Sussex Weidenfeld Institute of Jewish Studies] 2023