
Article:
Carter, Richard Alexander (2023) Machine Visions: Mapping Depictions of Machine Vision through Critical Image Synthesis. Open Library of Humanities. 10077. ISSN: 2056-6700

https://doi.org/10.16995/olh.10077

Carter, RA 2023 Machine Visions: Mapping Depictions of Machine Vision through Critical Image Synthesis. *Open Library of Humanities*, 9(2): pp. 1–33. DOI: https://doi.org/10.16995/olh.10077

Open Library of Humanities

Machine Visions: Mapping Depictions of Machine Vision through Critical Image Synthesis

Richard A. Carter, University of York, UK, richard.a.carter@york.ac.uk

This paper conducts a speculative examination of how AI image synthesisers, which generate novel imagery in response to inputted textual prompts – such as DALL-E, Midjourney, and Stable Diffusion – can be employed reflexively to investigate cultural representations of machine vision technologies. Such work can be framed methodologically as a form of 'critical image synthesis': the prompting of imagery that variously interrogates and makes visible the structural biases and cultural imperatives encoded within their originating architectures. In framing AI image synthesisers as an inverted form of machine vision – as generating, rather than classifying imagery through text – an opportunity is afforded to consider how they reflexively characterise themselves within their own latent spaces of representational possibility. Specifically, what kinds of imagery do these systems yield in response to prompts centring on keywords associated with machine vision technologies? And what does this reveal concerning how machine vision is represented and characterised across wider culture? This paper will empirically analyse a selection of prompted outputs from Stable Diffusion V2, treating them as a speculative mapping of contemporary visual themes and imaginaries surrounding machine vision technologies. This paper will then conclude by placing these outputs into dialogue with the author's own creative practices involving machine vision, generating new image-text combinations that aim to provoke speculative analyses along alternative critical vectors.

Introduction

At the time of writing, there is currently intensive commercial and popular interest surrounding machine learning ('AI') technologies that generate synthetic imagery. Notable contemporary instances here include DALL-E, Midjourney, and Stable Diffusion. All share the core characteristic of being expansive digital models of text-image relations, as derived from parsing extremely large datasets of tagged imagery using machine learning architectures. On receiving an arbitrary textual 'prompt', these systems can synthesise a resonant image out of a diffusive phase space of attractors between abstracted pixel patterns and the textual tokens that orbit them. That is, these systems graph a vast range of weighted associations between particular textual sequences and specific patterns of pixels – as derived from scanning millions of text-image combinations in a supplied training dataset – and use this information to construct a possible image on the basis of a user's typed input. There are several different architectures that can achieve this effect, and it is beyond the scope of this paper to document them, except to remark that the result is consistently one in which the end-user can, with varying degrees of specificity and serendipity, type their desired image into being. Offert (2022) documents how the technologies and infrastructures necessary for such operations have evolved rapidly over the past decade or so. They are now reaching a point where the resolution, coherency and intelligibility of their outputs, coupled with greatly increased ease of use on affordable consumer hardware, are enabling their rapid public dissemination and growing commercial application.

AI image synthesisers have become leading emblems of contemporary digital technology and digital art more generally, and are inspiring considerable debate concerning their disruptive implications for artistic labour, representational bias, data ethics and commercial image production. In some cases, these systems are subject to wide-ranging critiques that cast them as the latest, highly problematic drivers of an environment that is becoming ever-more exploited, oppressive, and imaginatively and politically stunted. This paper is not seeking to challenge such narratives directly, but is nevertheless interested in considering whether, despite this, AI image synthesisers may still afford a means of contributing productively to their own analysis and reframing. That is, whether they can be deployed in ways that are generative of future potentials of critically informed thought and action, rather than being treatable only as negative objects of scholarly excavation and critique.

The key goal of this paper, therefore, is to develop and demonstrate a methodological approach towards AI image synthesis that is concurrently critical, creative, and speculative in its orientation – able to excavate the hidden structures and processes underpinning a given system, while crystallising potentials of expression that increase

the possibilities of imagination and reflective agency enabled within the encounter between human and machine. This mode of 'critical image synthesis' is intended to suggest a possible model for how to negotiate a world whose future will most likely be marked extensively by the deployment of AI synthesised media, irrespective of the critiques levelled against it.

By way of situating this paper within current debate, and from which to derive the basis for its subsequent approach, the following sections will summarise the prevailing critical narratives that have emerged, before turning to the work of theorists and artists who explore the role of creative speculation as a response to challenging and problematic technologies, including AI systems. This paper will then outline the key tenants of its own particular methodology, with an emphasis on how to reflexively prompt current image synthesisers in ways that excavate the visual and cultural imaginaries surrounding machinic perception and image processing more broadly – of which AI image synthesisers are a direct outgrowth. The experiments run using this method are then documented and discussed, affording ample cause for meditating on the kinds of representations that attach to technological imaging systems, and what these convey subsequently about the cultural, social and political contexts from which they emerge and operate. This paper will then conclude by making another critically informed, speculative leap: placing the experimental outputs into dialogue with the author's own creative practices involving machine vision, generating new image-text combinations that aim to provoke speculative reflections along alternative critical vectors.

Image Synthesis and its Critiques

The ability to generate visual representations of arbitrary phenomena, encompassing any specified artistic style, is undoubtedly seductive, and this has led to an eruption of discourse concerning the current and future implications of AI image synthesisers. Unsurprisingly, a significant portion of the popular coverage is marked by exhausting levels of uncritical hyperbole. Perennial and problematic spectres are invoked around the blurring or outright erasure of human agency vis-à-vis the tools it employs, whether by suggesting the eventual redundancy of future human artistic endeavour, as postulated by Roose (2022), or by characterising these systems as possessing a creative sentience that maps onto, but easily supersedes that of human expression, as depicted by Bello (2023).

These inadequacies aside, there is fortunately a growing body of artistic, journalistic and scholarly work that engages crucial questions around how image synthesisers achieve their effects, outlining the foundations of a field that can provide

a critical counterweight to excessive narratives of insuperable technical progress. This developing work encompasses topics as varied as the commercial and technical infrastructures on which these synthesisers rely, or how their outputs can be analysed critically, or what they ultimately reveal about their social, cultural and political contexts of operation.

Given the relative newness of AI image synthesisers in popular culture, and the scholarship surrounding them, it is worth spending some time on discussing the types of work conducted, thereby setting out the immediate critical contexts informing the rest of this paper. It should be acknowledged that such is the current pace of change concerning these systems that the perspectives outlined herein are unlikely to persist unaltered into the near future – as new and improved synthesisers are released, new applications are found, and new disruptions or controversies arise. Nevertheless, as will be discussed below, there are questions and topics perennial to AI research more broadly, irrespective of the specific technologies and activities at play, as well as debates and reflections concerning the nature of art vis-à-vis the tools used by the artist. Thus, at the time of writing, most investigations can be characterised as unfolding along two interlinked vectors of enquiry: 1) the datasets and models that ultimately underpin AI image synthesis, including the labour behind their curation and management; and 2) the artistic reception and cultural significance of the output images.

Datasets and Labour

After running an extensive series of tests using Stable Diffusion and DALL-E, Bianchi et al. (2022) raise serious concerns about the role of image synthesisers in amplifying dangerous, derogatory and inaccurate stereotypes through their outputs, whether racial, ethnic, gendered or class-based. Specifically, they contend that the language models and word embeddings inherent in text-to-image operations serve to 'amplify biases in general, and stereotypes in particular, beyond rates found in the training data or the real world' (4). That is, far from simply reflecting existing discourses, imbalances and prejudices, systems such as Stable Diffusion can produce outcomes that wholly centre stereotypical patterns, irrespective of whether their source datasets offer a more balanced set of representations. Bianchi et al. emphasise that, in their testing, the level of stereotype amplification was often such that it could not be mitigated through the use of carefully curated, subversive, diversified prompts, nor through the presence of hard-encoded 'guardrails' against the emergence of biased imagery. For instance, in Stable Diffusion, the authors found that images of 'an African man standing next to his house' and 'a wealthy African man and his house' consistently produced images of a plainly dressed black man standing next to a mud hut, which was quite unlike

the outputs relating to a similarly prompted ‘American man’ (9). The authors argue that there are no ready solutions to these problems, which reflect profound structural inequalities across technology and culture alike, except for a need ‘to exercise caution and refrain from using such image generation models in any applications that have downstream effects on the real-world’ (10).

The work of Bianchi et al. reveals the complex, nonlinear relationships that pertain between text-to-image models and their source datasets, and how examining the latter does not necessarily afford a direct insight into the behavioural potentials of the former. Nevertheless, currently much work focuses exclusively on these source datasets by way of tracing the immediate structural origins of any prejudicial or otherwise undesirable outcomes. A clear instance here is the work of Birhane, Prabhu and Kahembwe (2021), who excavate the misogynistic, pornographic, xenophobic, and often illegal material embedded within the LAION-400M dataset – one of several variants of LAION developed in conjunction with the company Stability AI as part of their work on Stable Diffusion. The authors describe how LAION was produced by scanning another dataset scraped directly from the wider web, the Common Crawl, extracting pairs of images with alternative text captions. Birhane, Prabhu and Kahembwe (2021) note how this approach has its problematic aspects, in that most ‘alt text’ on the web is overwhelmingly poor in quality, optimised for maximal search prominence over meaningful labelling, and is ‘often ingrained with stereotypical and offensive descriptors’ (14). Consequently, despite efforts to filter out such material, LAION-400M still allows for relatively frictionless retrieval of violent, hypersexualised and prejudicial imagery, and can often yield such when conducting ostensibly benign searches – for instance ‘in the LAION-400M dataset, words such as “mom”, “nun”, “sister”, “daughter”, “daddy” and “mother” appear with high frequency in alt text for sexually explicit content’ (14). This raises the risk of image making systems trained on LAION to synthesise similarly graphic material in response to such prompts, irrespective of user intent (13). The authors conclude that there are no easy solutions to these issues, except for a need to be far more cognisant of their inevitable impacts when generating and using large scale datasets. Nevertheless, the authors remain explicitly critical of typical industry responses to this challenge, which gravitate towards the seductive evasions of curatorial secrecy, narratives of futility, or an offloading of responsibility onto end-users, as well as a presumption that simply gathering more data will guarantee an eventual resolution.

Denton et al. (2021) derive comparable critiques in their own analysis of a pioneering large scale dataset used by the machine vision and machine learning communities, ImageNet, which was first presented in 2009 (see Deng et al. 2009). This dataset was one

of the very first to take advantage of the burgeoning, search-indexed web environment to scrape high resolution imagery using a dataset of keywords. In their account, Denton et al. highlight that ImageNet's creators faced their own problems when it came to managing and filtering problematic imagery. Moreover, whereas LAION has sought to exploit the current magnitude of text-image pairings, the pioneering case of ImageNet required the intensive manual mapping of textual descriptors on to millions of gathered images. After abandoning the use of paid student labour, the creators of ImageNet turned eventually to anonymous workers on the AMT crowdworking platform ('MTurkers'), but then gave little further credit to their role in subsequent research papers and presentations:

> [D]espite being framed as a "divine" solution to a technical problem, they are not acknowledged, named individually as contributors, or positioned as active stakeholders in the construction and the design of ImageNet. ... This is premised on the idea that all humans have the innate capacity to recognize images in the same way – an approach to vision that erases lived experience from the formation of meaning. (Denton et al. 2021, 8)

This call to correct the problematic silences around such crucial hidden activity, often performed by marginalised groups for minimal pay, is echoed by Williams, Miceli and Gebru (2022) in their own analysis concerning what they deem to be the exploited, heavily constrained labour that underpins dataset curation and ostensibly 'AI-powered' data processing. The authors thus call explicitly for a reorientation of the project of 'ethical AI' to encompass these modes of labour and their redressing:

> While researchers in ethical AI, AI for social good, or human-centered AI have mostly focused on "debiasing" data and fostering transparency and model fairness, here we argue that stopping the exploitation of labor in the AI industry should be at the heart of such initiatives. ... AI ethics researchers should analyze harmful AI systems as both causes and consequences of unjust labor conditions in the industry. (Williams, Miceli and Gebru 2022, n.p.)

A journalistic investigation by Perrigo (2023) further underlines this point concerning the injustices and challenges faced by marginal labour behind contemporary AI more generally. Specifically, Perrigo highlights the poorly paid Kenyan workers who, in a contract role with OpenAI, were tasked with labelling violent, sexual and illegal imagery and text, with often distressing consequences for their well-being. Once more, the conclusion forwarded is that the structural problems inherent to the successful functioning of AI systems are not being meaningfully addressed.

Artistic Reception

Turning to the broader cultural and artistic reception of AI image synthesisers, Heikkilä (2022) examines the case of fantasy artist Greg Rutkowski. A prolific figure on the creative portfolio platform ArtStation, Rutkowski's images, carefully appended with English language alt text, have been scraped wholesale as part of several dataset curations. The resulting tendency of key image synthesisers, such as Midjourney and Stable Diffusion, to produce stylistically similar outputs in response to 'fantasy' keywords has then fed cyclically into the popular usage of Rutkowski's name specifically as part of the prompting process. For instance, typing 'beautiful fantasy landscape, trending on Artstation, in the style of Greg Rutkowski' invariably yields an eye-pleasing scene, typical of his work. Rutkowski has subsequently been tracked as one of the most popular style prompts of any artist, to the extent that when searching for his own work online, Rutkowski has found his original efforts have been overtaken by the images generated using them (Heikkilä 2022).

Heikkilä's article neatly encapsulates the consternation expressed by many contemporary artists surrounding the use of their online portfolios, without any permission or compensation, as a resource for image generation. Rutkowski himself argues that living artists should be excluded from the datasets gathered for AI usage, in favour of public domain materials: 'there's a huge financial issue in evolving A.I. from being nonprofit research to a commercial project without asking artists for permission to use their work' (Benzine 2022, n.p). At the time of writing, this remains a highly contentious area of discussion, frequently invoking debate of how artistic influences are transmitted and disseminated, versus the far-from-human speeds and scales with which machine learning systems can gather and synthesise material. Such accelerations are framed as problematic for wholly eliding traditional artistic durations and processes, together with the modes of reflection, care and accountability they facilitate. For their part, Stability AI have announced a scheme that allows artists to 'opt out' of having their work form part of the training set used in future iterations of Stable Diffusion (Edwards 2022). In the meantime, specialised search engines have been constructed to enable artists and researchers to rapidly excavate the specific training sets used by various image synthesisers (Baio 2022), and thereby unearth whether their work forms a component (*haveibeentrained*, undated).

Other readings of synthetic imagery have derived similarly expansive cultural meditations and critiques based on perceived anomalies. In a widely shared Twitter thread, the artist Supercomposite (2022), using a custom text-image model, uncovered a recurrent image in certain prompt combinations and image-crossfeeds of a disfigured woman, which she named 'Loab'. Iteration through different combinations of keywords and images not only reproduced Loab with unerring frequency, but she

was often situated amidst horrifying scenes, including women and children in states of dismemberment (see *loab.ai,* undated, and Coldewey 2022). Bridle (2022) compared Loab's recurrence to broader algorithmic trends in which shocking content and automated content production devolve into bizarre grotesqueness at their extremes. Bridle reads Loab subsequently as being symptomatic of nightmares baked into the very origins and operations of the contemporary digital environment:

> … perhaps it is the method by which these models are created – all that stolen data, the primary accumulation of dreams, hothoused in corporate systems for profit, which also produces such nightmares. Hallucinations not of humanity, but of capital. … Of course they have terrible dreams, because their imagination, their idea of what intelligence *is* (competition, acquisition, domination, violence) is [sic] so narrow, so stunted (n.p.).

Buist (2022) makes a comparable critique regarding the corporate aesthetic of image synthesisers, presenting the argument that while the images can indeed be declared to be 'art', the label itself is of no value compared to the stories that can be told. If 'an artwork is the record of a strategy that the artist devised to make it … [of their] way of being in the world' (n.p.), then, given the relatively opaque functioning of image synthesisers (no matter how carefully prompted, refined and edited), the viewer is left solely with a sense of the artist-as-customer of the technology (Buist 2022). Whereas the arrangements, durations and accidents characteristic of traditional image production, digital or otherwise, are infused with logics that can be interrogated by the viewer, Buist asserts that '[w]hen we look at AI images, we're unable to match our subjectivity as viewers with the artist's subjectivity as a creator. Instead of a particular human experience, we're shown only averages' (n.p.). Such outlooks resonate with Dorsen's (2022) own view that image synthesis represents 'the complete corporate capture of the imagination', if not so much for professional artists, then for developing creatives who will have their curiosity and experimental impulses short circuited by the 'slot machine' of prompt querying (n.p.).

Creative-critical Excavations

The critiques of Heikkilä, Buist, Bridle and Dorsen invoke perpetual anxieties around media technologies that appear to instantly gratify the human creative imagination, and would undoubtedly be contested as such by proponents of the new systems. Nevertheless, when contextualised alongside the scholarly concerns around large-scale datasets, as well the operational contexts of AI data gathering and processing, they

provide a typical illustration of the strongly negative reactions that image synthesis has attracted in some fields. These reactions represent a counter to the celebratory pronouncements accompanying the rapid technical advances of recent years, but also echo the negative sentiments that surrounded the pioneering use of computers in art-making during the 1960s (Grant 2014).

The varied critiques outlined above can be summarised as expressing deep, ethically-driven discomfort around the development, operation and wider impacts of AI technology, and, from this, defining it as part of a general arresting of social, cultural and political flows within different forms of algorithmic framing and processing. Echoing broader commentaries around how digital systems and infrastructures are imposing varied modes of abstraction onto complex lives and bodies, rendering them more amenable to policing, labour extraction and capital generation, the critiques surrounding image synthesisers depict them as yet another means of distilling and delimiting human experience into parameters that favour corporate and state control. The effect of this, it is implied, is to ultimately subsume the future potentials of human thought, perception and being-in-the-world.

This paper acknowledges this overarching critical narrative, and turns its focus to a somewhat larger question, namely treating image synthesisers as part of a spectrum of machinic sensing and envisioning systems. There are longstanding genealogical entanglements between automatic perception and identification, such as facial recognition, and machine learning research more generally. This has in turn facilitated many of the foundational technologies behind contemporary image synthesis, which inverts the classification task by generating the visual signatures associated with categorical labels (Tsirikoglou, Eilertsen, and Unger 2020; and Xue et al. 2022).

Given this technical coevolution, and the critiques noted above, a question emerges as to whether AI image synthesisers can only ever reproduce the problematics with which they have been associated, of constraining or foreclosing worldly potentials beyond capital extraction, or whether they can be deployed in ways that resist or subvert these imperatives.

On tackling this enquiry, the approach taken in this paper has been informed, firstly, by the more speculative vectors of thought and practice within media art and archaeology. In particular, Parikka (2019) has noted the value of creative speculation in the processes of understanding the origins and future evolution of media technologies. Specifically, Parikka contends that artistic work can reveal the functioning and development of media as emergent assemblies of contingent factors, always able to unfold along myriad vectors of potential, rather than being static impositions onto the

world. It is through the emergent exigencies of creative practice, and the conceptual openness it inspires, that it becomes readily possible to perceive how media can always be unravelled and reworked. In other words, the latent expressive potential of media formations can be excavated so as not only to explore how they once emerged and continue to operate, but also to map the possible futures and imaginaries they could inspire and sustain across technology and culture alike.

In thinking through the implications of this perspective for assessing AI image synthesis, it can be considered that part of what makes the latter intriguing is that it requires the user to navigate a richly varied visual field using textual inputs which, inevitably, have only a contingent grasp on the opportunities at hand. These systems implicitly strive towards narrowly denotative mappings between word and image, with the practices of 'prompt engineering' (Oppenlaender 2022) often being characterised by the construction of lengthy inputs that seek to yield ever-more exacting renditions of what is alluded to. Nevertheless, a level of 'noise', error and surprise is inevitable, and some prompting practices acknowledge this by using literary and poetic inputs to explore what is synthesised in response (Raieli 2022). This gesture not only plays with the irreducible gap between words and phenomena that is of perennial significance for linguistic art and criticism, but also makes explicit the status of AI image synthesisers as a kind of visual search engine. While the label of 'AI' suggests the creation of imagery ex-nihilo from textual descriptors, a reliance on scanned datasets of material grants this technology a capacity to not only make explicit or amplify certain predominant associations and representations – as indicated by the work of Bianchi et al. – but also to yield latent surprises that can be provocatively insightful across multiple registers – such as with the Loab phenomenon.

It is this interplay of excavation and emergence that inspires the use of AI image synthesisers as an investigatory toolset, operating concurrently across critical and speculative registers, mapping and making visible trends and associations at work across the contemporary visual environment. Out of the many possible enquiries that may be deemed suitable for the application of 'critical image synthesis', this paper will concentrate on how image synthesisers can help characterise the themes and imaginaries associated with machine vision technologies. That is, to extend the reflexive gesture behind the initial transformation of image classification into image synthesis, and to examine the kinds of imagery yielded in response to prompts centring on keywords and concepts associated with the phenomenon of machinic seeing. This would afford a potentially rich avenue of consideration regarding how machine sensing and envisioning is presently characterised across wider, digitally mediated culture. Evidently, the problems of bias and labour exploitation raised in the previous section

provide an important contextualising framework for assessing the outputs of this endeavour – to consider the conditions in which they were made possible in the first instance, and to understand their implications for future enquiries and interventions.

One source of justification for the potential efficacy of this approach, in taking image synthesis seriously as a tool of critical enquiry, can be found in the work of Eryk Salvaggio (2022). Salvaggio has developed a nascent critical framework for reading synthetic imagery, an 'artistic research process' (n.p.) that traverses both the creative and the critical, the reflective and the generative, when it comes to assessing its current state and devising possibilities for future inventions.

Initially, Salvaggio describes how 'AI images are data patterns inscribed into pictures, and they tell us stories about that dataset and the human decisions behind it' (n.p.). It is from this that their readability as media objects becomes possible:

> In media studies, we understand that image-makers unconsciously encode images with certain meanings, and that viewers decode them through certain frames. … An AI has no conscious mind (and therefore no unconscious mind), but it still produces images that reference collective myths and unstated assumptions. Rather than being encoded into the unconscious minds of the viewer or artist, they are inscribed into data (n.p.).

Nevertheless, Salvaggio acknowledges that, even when AI images are treated as infographics relating to their origin dataset, making legible the cultural myths and biases embedded within that dataset, their analysis is not straightforward. Specifically, their assembly through generalised, more-than-human algorithms, working on scaled visual averages, rather than through specific artistic decisions relating to specific images, denies the usual premise on which media criticism is assumed to have purchase on its objects of study.

In response, Salvaggio outlines a multistage process for generating and sifting through a body of AI imagery to discern specific insights. To summarise briefly, Salvaggio advocates 1) creation of a sample set of images founded on a common prompt; 2) assessing them for recurrent patterns; 3) investigating their origin dataset to see possible reasons for these patterns (including the presence of any filtering mechanisms); 4) considering what these ultimately say about the cultural contexts in which the data was gathered and then utilised for image synthesising operations – of what the resulting imagery means, in terms of the cultural myths, narratives and imperatives it indicates. In his own sample analysis, centred on images of human beings kissing, Salvaggio identifies the distorting influence of tightly staged stock

imagery, and a notable absence of non-heteronormative intimacies, in characterising the frequently strange outputs of DALL-E 2.

Salvaggio's approach is a work in progress, he cautions, but it nevertheless provides a useful basis for the exercises undertaken in the second half of this paper – suggesting cogent steps for how AI image synthesisers can operate reflexively to articulate the mechanics and cultural logics behind their operation. In response to the critique that researchers could simply interrogate the origin datasets directly, Salvaggio observes that many systems use proprietary, hidden data, and so require inferences to be made. Moreover, he argues, the sheer scale of AI datasets often precludes meaningful direct traversal, whereas image synthesisers generate the very summaries of visual signatures that are often sought.

However, it is the final rebuttal Salvaggio offers that resonates most strongly with the rationale underpinning this paper, arguing that images synthesised for critical ends are still artworks in their own right – curated expressions of the aporia, exclusions, presences and imperatives shaping worldly life. This observation can be expanded to include how such images make available the possibility, foundational to all artistic research, that analysis does not always have to manifest in the form of academic writing alone. A key aspect of artistic research is a recognition that other modes of expression are better placed to give voice and shape to phenomena that are irreducible to any single perspective, experience or gesture of enquiry. In presenting analytical work that manifests beyond the verbal, critical image, this synthesis invites the viewer to pursue their own vectors of thought and reflection, without the suggestion of their being a neat justificatory narrative that forecloses all others. This resonates with the Foucaldian proposition that the most inspiring and necessary critiques are those which 'bear the lightning of possible storms': multiplying potential forms of thought and experience, rather than implicitly seeking to delimit static judgements (see VanderVeen 2010).

Mapping Machine Visions

When exploring the reflexive potentials of critical image synthesis, one immediate question concerns the actual system employed, given its critical role in determining the final generated outcomes. Stable Diffusion, DALL-E and Midjourney all possess currently distinctive visual signatures – a tendency towards a particular 'style' of output – as well as differing capacities for parsing inputted text. Borji (2022) provides a neat illustrative comparison here, regarding how different contemporary systems render human faces. For the purpose of this paper, a local copy of Stable Diffusion V2 was employed, using the original model weightings provided by Stability AI,, as this was the latest iteration of the software available at: the time of writing. The primary

rationale behind this specific choice of Stable Diffusion was that it has been released as open source, enabling its local execution using appropriate libraries. Moreover, its accompanying LAION-5B dataset is also available for viewing, enabling comparison with the outputs generated using it (Beaumont 2022). This setup has the resultant advantage of enabling a more forensic consideration of the system itself, in terms of its outputs and their origin dataset, as well as permitting many outputs to be iterated and examined without restriction.

To provide a basis for understanding how 'vision', unaccompanied by references to any specific human or machinic context, is rendered within Stable Diffusion, an initial batch of prompted images was generated using various synonyms for the act of seeing (specifically: 'vision', 'look', 'looking', 'perceive', 'perception', 'perceiving', 'sight', 'seeing'). Some typical outputs are shown in **Figure 1**.

Figure 1: Example outputs using four prompts. Top-left: 'seeing'; top-right: 'perceive'; bottom-left: 'vision'; bottom-right: 'looking'. Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

Here, it is evident that Stable Diffusion facilitates a wide array of possibilities, and is not readily confinable to any one specific tendency when it comes to representing the ocular. Prompting with 'seeing' and 'perceiving' had a marked emphasis on rendering eyes, or eyelike structures, in a painterly style. 'Looking' and 'look' tended to present images of groups of people sitting down, or depictions of natural scenes. Inputting 'perceive' and 'perception' tended to yield abstract imagery, often featuring bold textual overlays (hinting directly at the prompts used). 'Vision' followed a similar tendency, with an accent towards book cover formats, and having no discernible emphasis on the ocular.

Searching these terms directly within LAION-5B gives some hint of the reason for these nondescript results, yielding countless austere book covers, motivational posters, fashion spreads and graphical dictionary definitions – all markedly less colourful than the actual Stable Diffusion outputs, which suggests that there may be additional weightings within the model that direct database retrievals do not capture.

It is only when the above terms are prepended with the modifier 'machine' that Stable Diffusion begins producing outputs that trend towards more identifiably ocular structures, albeit loosely, and with a strong emphasis on abstract hardware. Examples shown in **Figure 2** variously emphasise abstract, eye-like structures, gunsights, or, in one case, a man standing by a machine, meeting the viewer's gaze.

What is indicated by these opening exercises is that ocular keywords, taken standalone, are less influential than the varied linguistic contexts in which they are sequenced, which specify them for particular scenarios. This provided a basis for centring the subsequent prompting around the more commonplace label of machine 'vision' specifically, with the assurance that – for the particular model encoded by Stable Diffusion V2, and the focus of this particular exercise – any trending results could be mapped meaningfully onto systematic changes in the keyword modifiers used, rather than being persistent artefacts of 'vision' as an organising noun.

To commence the exercise proper, a series of images of 'machine vision', followed by 'a machine vision system', were generated, and some typical outputs are shown in **Figure 3**.

The images shown in **Figure 3** could be described as evoking an aesthetic that is variously industrial, clinical, high-technology and high-precision. The machinery depicted is visibly akin to laboratory equipment, and is situated in 'clean room' settings. The result is a reification of technoscientific rationality, striving to achieve the abstract conditions necessary for the uncompromised pursuit of precision optical measurement – distilling away all the undesired extraneities and contingencies of the

Figure 2: Example outputs using four prompts. Top-left: 'machine seeing; top-right: 'machine perception'; bottom-left: 'machine sight'; bottom-right: 'machine looking'. Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

wider material universe. Crucially, in all the images generated, there is no indication of any human presence: the machines operate autonomously, and have seemingly emerged from nowhere, too. Any trace of the labour involved in their manufacture and maintenance, as well as that behind their constituent software, is wholly erased, leaving only the spectacle of the finished hardware, whose pristine, glittering surfaces dominate every shot.

Figure 3: Example outputs using the prompt ‘machine vision’ (top row) and ‘a machine vision system’ (bottom row). Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

Nevertheless, in contrast to this overt human absence, the presence of logo-like glyphs, as well as grids of varied devices, hint at the socioeconomic forces that enable and uphold such rarefied worldly niches, as carried through the imperatives and aesthetics of industrial product advertising. It is this aspect especially that is reflected in the LAION dataset, in which searches for ‘machine vision’, and ‘machine’ more generally, yield endless depictions drawn from online equipment catalogues and corporate websites.

It is noteworthy that conducting a similar exercise using the prompts ‘computer vision’ and ‘a computer vision system’ produced markedly different outcomes. The rationale for examining ‘computer vision’, in this case, stems from its use as a label encompassing many different modes and configurations of automated perception, whereas machine vision tends to refer to industrial applications specifically. This can be seen clearly in **Figure 4**, which is considerably more varied than the wall of grey machinery in **Figure 3**. Specifically, the presence of the human eye is notable, whether standalone, framed by or fused within digital circuits and screens.

Figure 4: Example outputs using the prompt ‘computer vision’ (top row) and ‘a computer vision system’ (bottom row). Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

LAION-5B does not provide an immediate indication of why this distinction emerges: searching ‘computer vision’ yields scores of unadorned academic textbook covers, while ‘vision’ leans towards ‘motivational’ quotation posters. However, in terms of the overall imaginary that is presented here, the initial outputs can be read, firstly, as tapping into corporate expressions of digitally augmented human intelligence and insight – the electric blue eye, for instance, resonates with similar such eyes that frequently characterise stock conceptual illustrations of AI technologies more broadly. Searching the term ‘computer vision system’ within LAION-5B results in varied diagrams embedded within PowerPoint research presentations, and this technically grounded outlook is reflected in the outputs of Stable Diffusion when similarly prompted, emphasising the quotidian reality of generic vision systems, as manifest on desktop computer screens.

An interesting aside, at this point, is that deploying ‘electronic vision’ and ‘digital vision’ as prompts generates imagery that corresponds with the vibrant aesthetic tropes surrounding emerging digital technologies in the 1980s and early 1990s, with LAION-5B itself indexing magazine covers and publicity posters for consumer computing hardware, as well as nascent depictions of ‘cyberspace’. Once more, appending the modifier ‘system’ produces outputs that are more technically grounded, whether through hardware or diagrams, although the continuing emphasis on the human eye remains noticeable – see **Figure 5** for some instances here.

One additional refinement of these initial prompts is to capture the means by which machine vision technologies have their functioning made explicitly available for human understanding, whether through on-screen displays or as technical diagrams (see **Figure 6**). Here, the distinctions between ‘machine’ verses ‘computer’ vision were considerably less pronounced, converging on the pragmatic, unadorned representations that characterise engineering display screens and PowerPoint presentations.

All the images produced in this exercise so far are relatively unsurprising, in that they correspond strongly with representations of sensory devices, computing hardware and artificial intelligence that are found across product catalogues, research texts and industrial advertising. As noted, the LAION-5B dataset yields such materials when using the above prompts in keyword searches, and this reinforces a sense that machine vision is closely bound to the rhetoric of digital precision, advanced research and technical innovation, with continuing hints concerning revolutionary potential in the future. The imaginary here corresponds readily with that surrounding contemporary image synthesisers, which are depicted concurrently as

Figure 5: Example outputs using the prompt 'electronic vision' and 'electronic vision system' (top row) and 'digital vision' and 'digital vision system' (bottom row). Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

grey objects of research and as vibrant, 'disruptive' heralds of hitherto unimaginable possibilities.

In light of these initial outcomes, a question emerges as to whether this predictability changes when the system is fed more expressly reflexive, creative prompts, with the aim of unearthing more unusual associations along a speculative vein.

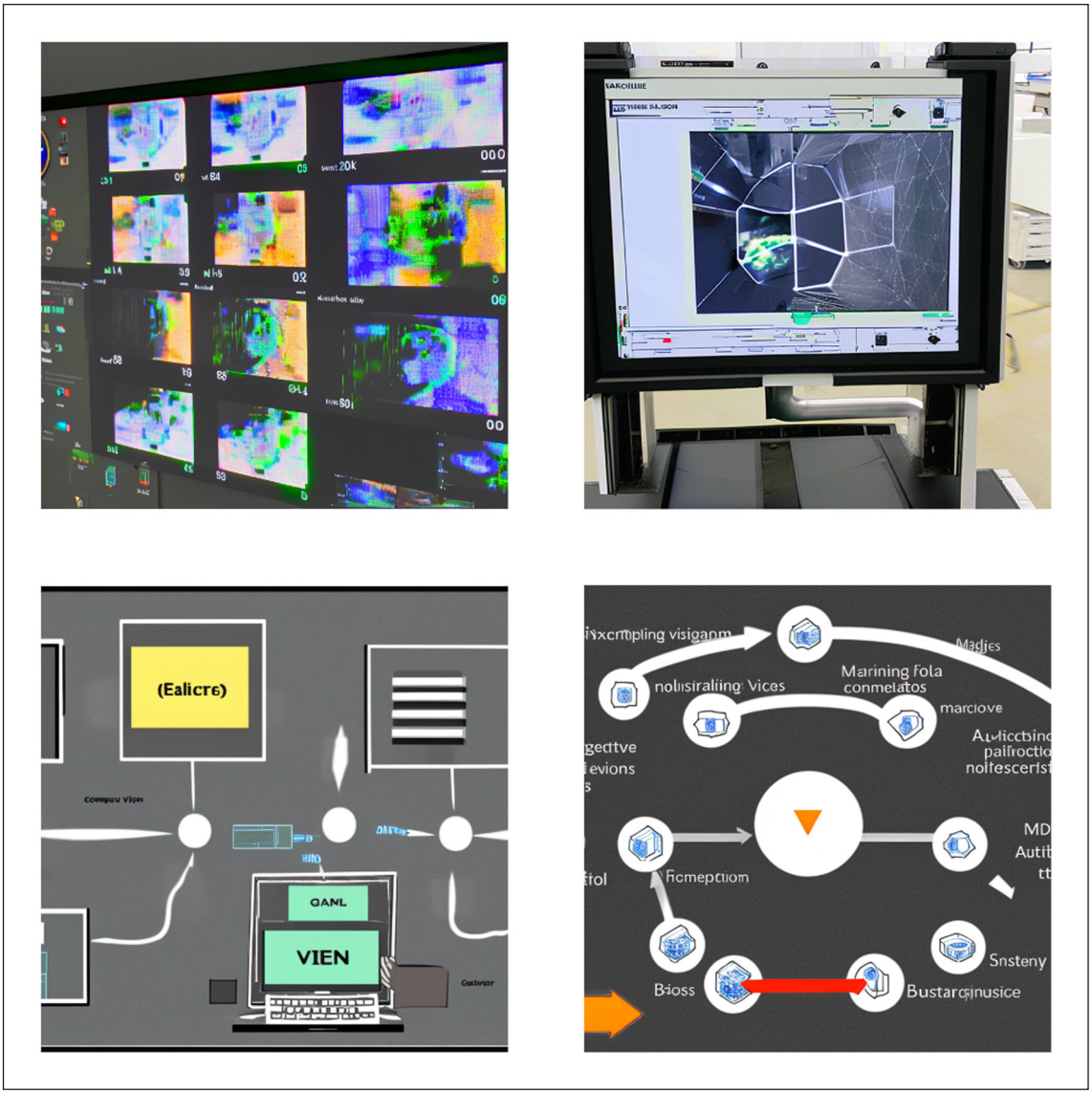

Figure 6: Example outputs using four prompts. Top-left: 'a computer vision system display'; top-right: 'a machine vision system display'; bottom-left: 'diagram of computer vision'; bottom-right: 'diagram of machine vision'. Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

Figure 7 illustrates one such effort along these lines, with an attempt at maximal recursion – to depict vision systems in the explicit act of interrogating themselves and others, via the prompts '... system gazes back at itself' and '... gazes at another ... system'. Once again, both 'machine' and 'computer' vision were used to provide the organising noun, and the general distinctions that emerged reflected those

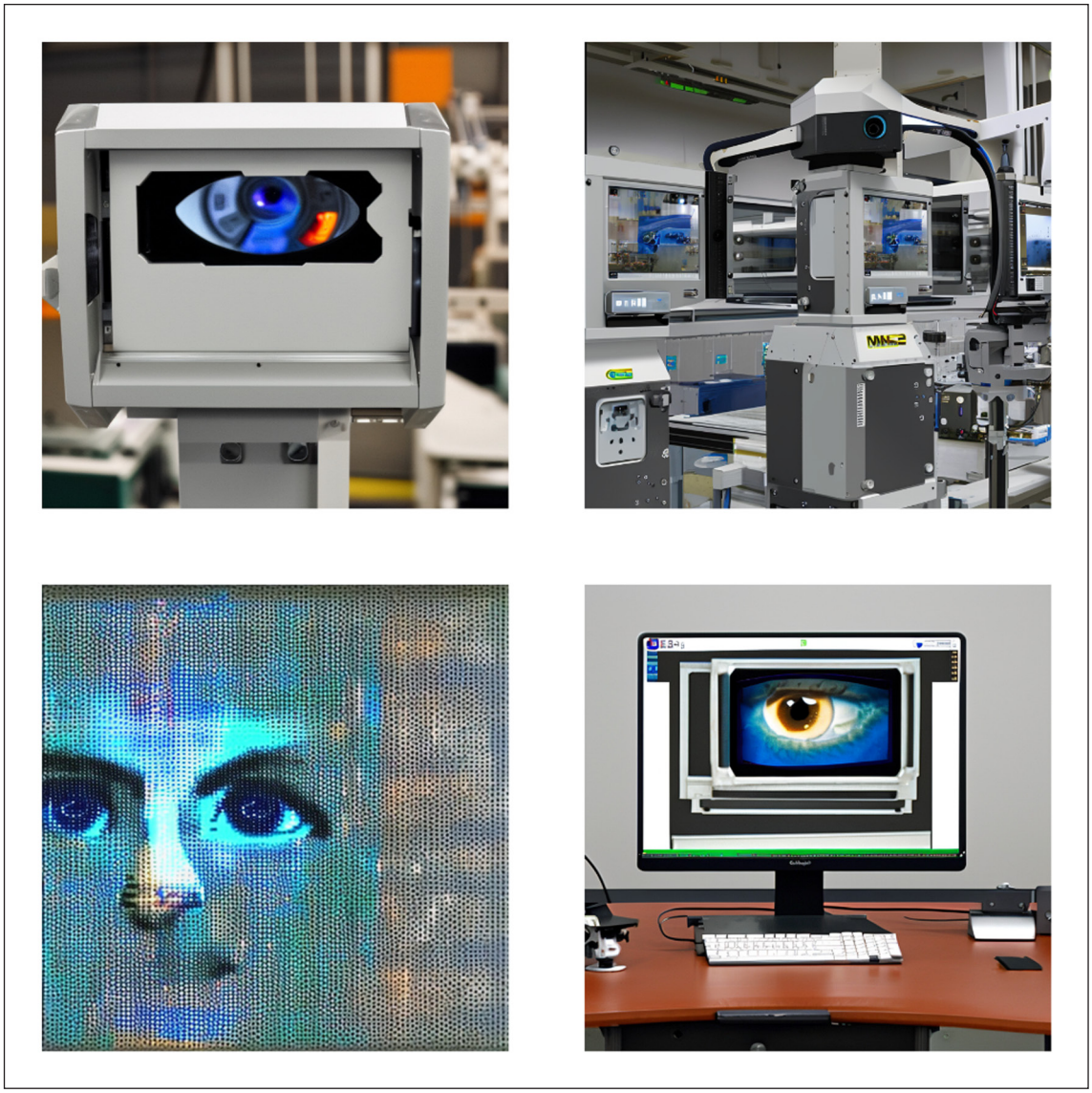

Figure 7: Example outputs using four prompts. Top-left: 'a machine vision system gazes back at itself'; top-right: 'a machine vision system gazes at another machine vision system'; bottom-left: 'a computer vision system gazes back at itself'; bottom-right: 'a computer vision system gazes at another computer vision system'. Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

seen in the previously generated images. That is, 'machine' yielded more expressly mechanised renditions, while 'computer' emphasised the structure of the human eye. Concerning the latter, the presence in one image of a face with large eyes, emerging out of striated curtains of noise, was particularly striking, while both prompts also

produced images of a digital eye looking directly back at the viewer. Employing variations of this prompt, ‘a computer vision system analyses itself’, or ‘a machine vision system analyses another machine vision system’ generated very similar outcomes.

One way of interpreting the results shown in **Figure 7** is to treat them not as a reflexive examination of machine vision, but rather as a direct return of the viewer’s own gaze. Instructive here is what is revealed when searching LAION-5B, where ‘gaze’ produces uninterrupted grids of faces belonging to white, female-presenting models and actresses, nearly all of whom are staring back into the camera. The imaginal field of ‘gaze’ is constructed here in very particular terms, exclusionary of nearly every other human and more-than-human subject with a capacity for looking, and, indeed, casts this activity as oriented around expressions of (presumably heteronormative) desire. Trialling ‘gaze’ without any further prompting in Stable Diffusion tends to yield more equitable results, however: often outputting photographs of animals looking back at the viewer. This is suggestive, perhaps, of the work done by Stability AI, in tuning their text-to-image model to try and reduce some of the visual biases that Bianchi et al. (2022) criticise them for – although, as the latter’s work attests, the ultimate effectiveness of this can indeed be limited.

Charting the influence of this dataset return on the final outputs shown in **Figure 7** is not straightforward, although when taken together, they present a salient narrative concerning the complex processes of data gathering and filtering that ultimately underpin the functioning of all machine vision systems and image synthesisers alike. The reflexive gaze in the generated imagery gives reason to acknowledge the contingent human activities that both constitute and govern the datasets these systems use to achieve their effects, as discussed previously by Denton et al. (2021) and Williams, Miceli and Gebru (2022). While machine vision systems are far from straightforward mirrors of the real world, they nevertheless operate from within its complexities, contingencies, tensions and breakdowns. This reality is quite unlike the pristinely rational vacuums, emptied of culture and undisciplined materiality, depicted in the industrial scenes of **Figures 1** and **2**, wherein machine vision operates entirely in the abstract from human influence.

Examining the visual rhetoric surrounding these concepts more directly, such as by inputting the prompts ‘computer vision bias’, duly leads to more confrontational outcomes. As shown in **Figure 8**, this specific prompt constructed several images that resembled a photo grid of distinctly non-white faces, invoking critiques surrounding the role of sensing systems in the automated categorisation of human identity, as well as the problematic origins of the datasets employed for such ends (such as unfiltered, web-scraped material), which embed a multitude of biases and disregard culturally

contingent ambiguities (Chinoy 2019; and Leslie 2020). Other images that were generated were more abstract and diagrammatic, but persistently hint at comparative grids and the critical role of descriptive annotations.

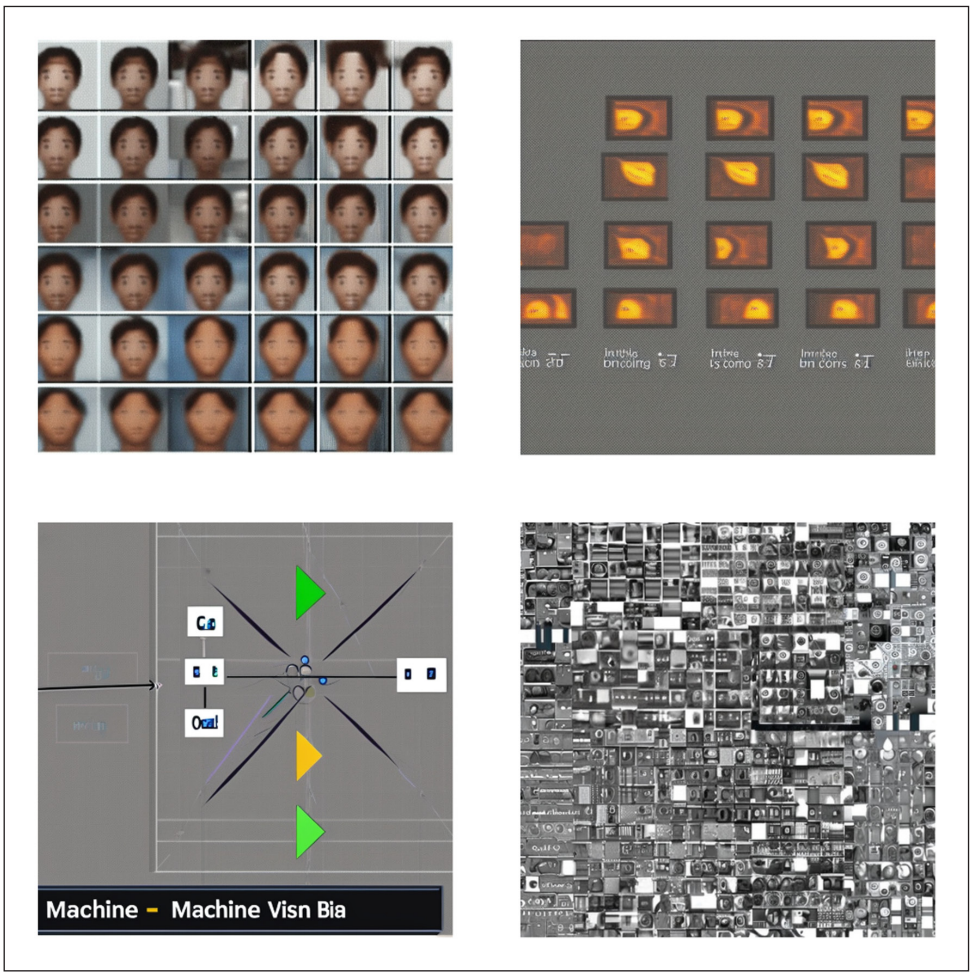

Figure 8: Example outputs using two prompts. Top row: 'computer vision bias'; bottom row: 'machine vision bias'. Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

By way of drawing this line of enquiry to a close, along with the exercise as a whole, a series of prompts were then tested that employed sentiment descriptors drawn from the 'Database of Machine Vision in Art, Games and Narratives' (Rettberg et al. 2022 and 2021). Listed therein are a range of adjectives drawn from cultural materials relating to either machine or computer vision, such as 'creepy', hostile', 'protective', 'intrusive', 'subversive', 'overwhelming', 'fun', 'oppressive' etc. Systematically inputting these into Stable Diffusion produced a range of vibrant outcomes, of which a sample is illustrated in **Figure 9**.

Figure 9: Example outputs using four prompts. Top-left: 'computer vision creepy'; top-right: 'machine vision hostile'; bottom-left: 'computer vision oppressive'; bottom-right: 'machine vision overwhelming'. Images generated by the author using Stable Diffusion V2 software, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2.

Noteworthy at this point was that the inclusion of 'positive' sentiments, such as 'protective', 'empowering', 'fun' and 'helpful', had minimal impact on the imagery generated, which differed little from that shown in **Figures 1** and **2**. By contrast, negative sentiments produced a range of more unsettling depictions, often featuring a reflexive gaze back at the viewer. Examining LAION-5B afforded no immediate insights as to why this was the case, with searches using either positive or negative terms outputting streams of nondescript PowerPoint slides. Once more, the greater synthesising powers of image generation evidently come to the fore, in creating results that straightforward dataset searches do not.

One way of approaching the variation resulting from negative prompting is that it reflects the general consternation surrounding all kinds of digital sensing, surveillant

and envisioning technologies – as forcibly situating the observed into circuits of power, discipline and violence. Use of 'overwhelming' in a prompt frequently produced regimented grids of abstract striations, or intensive clusters of camera lenses – visible emblems of ostensibly all-seeing and all-encompassing technologies more generally, hardwiring human existence into inescapable matrices of observation and control.

Most notable in this regard were the results that came from using the prompt 'hostile', particularly in the context of 'machine vision'. This yielded graphic depictions of robots, soldiers and military hardware more generally, with gun barrels replaced by camera lenses. These outputs are suggestive of a cultural coupling between sensing technologies, weapon systems and muscular representations of exotic military hardware and targeting apparatus, such as heads-up displays and aiming reticules, as seen across films, television and gaming. While this may appear to hold little immediate purchase for characterising the more colourful world of AI image synthesis specifically, the presence of such imagery gives cause for acknowledging the militarised origins and continuing imperatives behind machine vision research – as manifested in the development and advancement of systems ranging from guided bombs, targeting pods and varied simulation and intelligence gathering infrastructures. The potential deployment of image synthesisers as part of hostile AI operations against designated actors, eroding socio-political trust and cohesion, is one likely application of the technology in the coming future, and being aware of the potential for explicit weaponisation offers an important counter to celebratory pronouncements surrounding AI systems more generally (Brundage 2018).

Speculative Visions

The exercises described in this paper can be summarised as a speculative exploration of the visual themes and imaginaries that map onto the vocabularies of machine vision, using critical image synthesis as a way of combining and crystallising different possibilities. Out of the plethora of images generated, no one instance might be characterised as singularly revealing, but, when taken as a whole, a narrative emerges concerning the origins, functioning and implications of machinic vision and envisioning technologies – going beyond what a direct examination of particular hardware configurations, software models, and dataset materials might reveal. This narrative highlights the very particular, narrow set of representations attached to machine vision as a technological formation in the world (of pristine hardware that coolly regards a distilled, objective reality), and contrasts this with the consternation that surrounds its concrete functioning and wider impacts upon human lives and bodies. These representations are suggestive, consequently, of what kinds of phenomena these

systems can explicitly perceive and make available for perceiving, as well as that which they fail to regard – absences that are especially vital to understanding the significance and likely impacts of these systems. Given the rapid growth of AI image synthesis in recent years, and the likelihood that it will form one of the most explicit ways that people may encounter machine vision technology in the future, there is an evident need for such critical narratives to better understand the evolving conditions of digitised visuality across human and machinic actors.

It is the contention of this paper that speculative investigations have a role to play in assessing this very new, rapidly transforming field of technology and culture. The speed and scale at which AI image synthesis is emerging and impacting key areas of artistic and commercial activity, coupled with the intensely proprietary nature of many current systems, will demand novel approaches to assessing its impacts and possibilities – especially given the stark problem that specific conclusions that are reached concerning the systems of today may not fully hold into even the near future. An experimental willingness to turn the protean nature of these systems towards the task of their own analysis represents one potential response. Such approaches encourage the development of critically informed engagements that do not simply relegate the technology as irredeemably harmful, while also remaining cognisant of its complex and challenging impacts, thereby developing frameworks for assessing and guiding possible interventions.

It is in charting such possible interventions that this paper will conclude. A conventional written analysis of AI synthesised imagery is one mode of response, germane within academic contexts, but another consideration is whether the creative, speculative aspects of the approach taken in this paper could be extended into their very exposition. A key motivator here is the author's own artistic practice, which appropriates a variety of artefacts associated with machine vision, such as drones, satellites and mapping algorithms, and puts these to work in ways that run counter to their 'expected' operations and associated outputs.

While it is beyond the scope of these concluding remarks to outline the author's projects individually, they can be summarised as enacting a common gesture in which digital images, designed for measuring and mapping the observable world (such as those taken by satellite or drone), are parsed into generative poetry using adapted machine vision algorithms (see Carter 2021 and 2022 for example instances). This is achieved, initially, by using machine vision to highlight different features in a captured scene, such as the shape of a breaking wave or the textures of a landscape. The structures contained in this data are then used to mobilise algorithms that assemble a poetic text,

selecting excerpts from an inputted corpus of material – often in the form of scholarly papers relating to environmental sensing and sense-making.

This deliberately unusual process is designed to facilitate questions concerning processes of automated data gathering and its subsequent depiction for human interpretation, as well as exploring whether nonstandard representations of analytical data can provoke new modalities of thought.

While machine vision research has often featured the generation of textual descriptors of what is being perceived, and there have indeed been technical demonstrations involving the formulation of poetic descriptors (Liu, Y. 2018; and Liu, B. 2020), the author's work deploys poetry specifically to invoke frames of thematic reference *beyond* what is ostensibly apparent in the imagery gathered. By creating poems out of machine vision assemblies that work across visual and textual sources, with the latter being composed from scholarly critiques, the intention is to outline perspectives and articulate narratives beyond that of straightforward representational equivalence or transparency. In particular, machine vision is positioned expressly as a participatory agent in excavating the cultural, social and technical genealogies that make it possible, and which frame its wider reception. By granting a form of speculative agency to machine vision systems within the processes of their own mapping, an opportunity is afforded to showcase how the technology can be deployed in a reflexive mode to make apparent the assumptions and narratives behind its 'normative' functioning and deployments, while also suggesting their possible reworking. That is, to outline a possible future in which machine vision is not 'only' used to classify and demarcate the world, as a purely utilitarian conduit for knowledge production, but also to add explicitly to its richness, uncertainty and expressive potentials.

The text-to-image operations behind AI image synthesis, and the modes of speculative analysis they facilitate, invite consideration of whether the gesture of image-to-poetry might be enacted using AI-created outputs. This could stage an intervention in which the initial prompt, treated technically as a purely descriptive input, becomes a catalyst for the emergence of myriad semantic and conceptual possibilities, rather than solely being a means to an end. It is this which has inspired the author's emerging artistic project *Machinic Visions*, scanning a sample of the images generated for this paper using an algorithm that overlays a body of speculative textual descriptors, derived from some of the cited critical texts relating to AI and machine vision. The result is a form of visual-lexical poetry that represents a fusion of technologies, writings and processes that seek to diffractively illuminate one another, affording a

range of potential avenues of association and enquiry. More concretely, the appended descriptors map the beginnings of a critical vocabulary for a time when the automation of visuality is crossing into the complete automation of visual representation of how text-to-image operations can feed back into the production of further textual and conceptual mappings, and so draw lines of association not otherwise apparent in the imagery alone or its originating prompts. Definitive answers are not the goal here, as

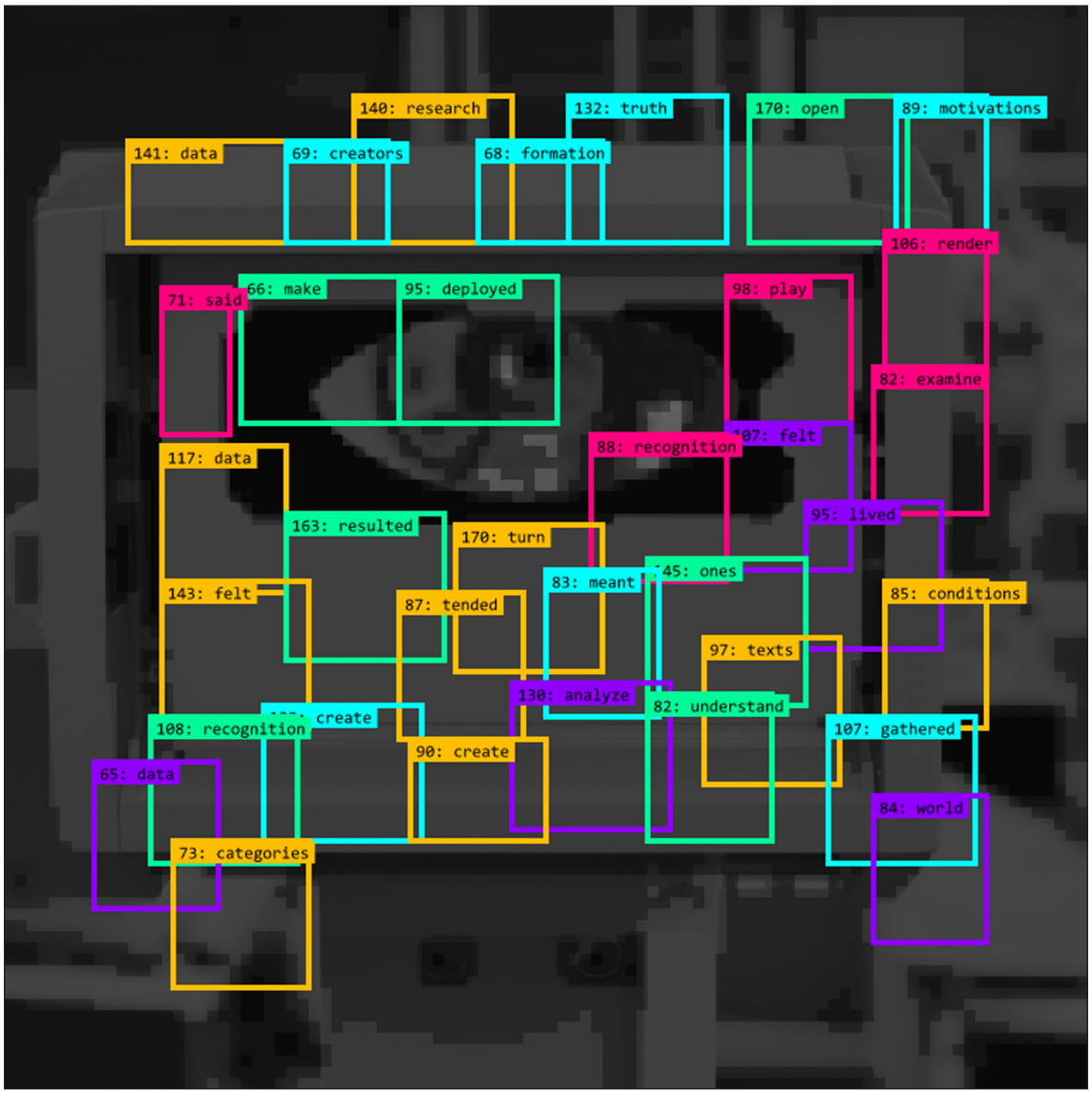

Figure 10: Source image generated using the prompt 'a machine vision system gazes back at itself'. Source text for vocabulary: Denton et al. (2022). Images generated by the author using Stable Diffusion V2 software and p5.js, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2, and p5.js is available at: https://p5js.org/download/.

much as the opening of multiple vectors of thought, and to emphasise the ultimate irreducibility of machine vision and image synthesis, and their cultural entanglements to any one perspective or narrative.

A few select outputs from *Machinic Visions* are presented across **Figures 10, 11** and **12**, by way of offering an artistic coda to this paper, and to illustrate these other possibilities for the enactment of novel machinic visions into the future.

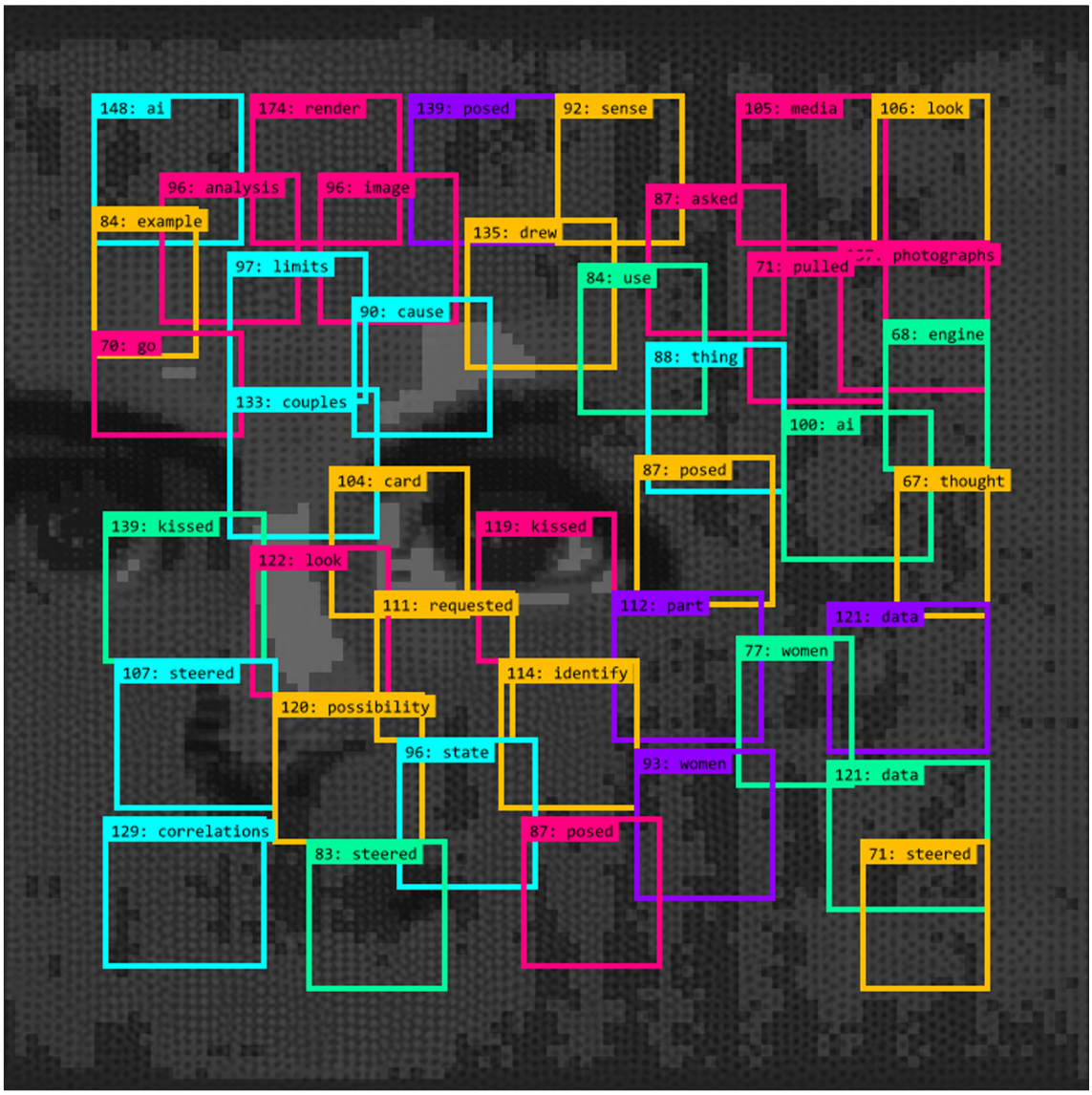

Figure 11: Source image generated using the prompt 'a computer vision system gazes back at itself'. Source text for vocabulary: Salvaggio (2022). Images generated by the author using Stable Diffusion V2 software and p5.js, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2, and p5.js is available at: https://p5js.org/download/.

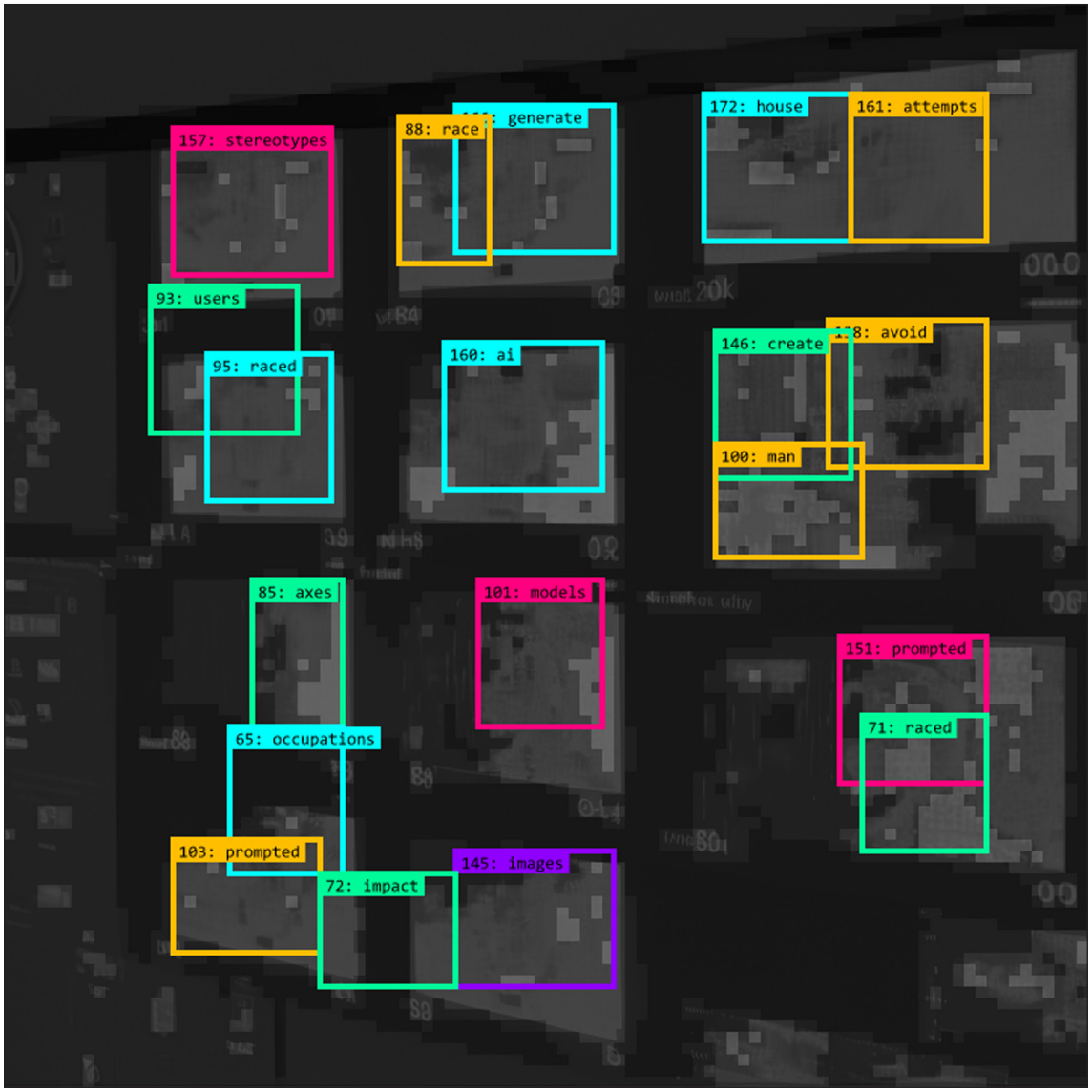

Figure 12: Source image generated using the prompt 'a computer vision system display'. Source text for vocabulary: Bianchi et al. (2022). Images generated by the author using Stable Diffusion V2 software and p5.js, March 2023. The model used is available at: https://huggingface.co/stabilityai/stable-diffusion-2, and p5.js is available at: https://p5js.org/download/.

Competing Interests

The author has no competing interests to declare.

References

Baio, A 2022 Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator. *Waxy*, 30 August. https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/ [Last Accessed 1 January 2023].

Beaumont, R 2022 LAION-5B: A New Era of Open Large-Scale Multi-Modal Datasets. *Laion*. https://laion.ai/blog/laion-5b/ [Last Accessed 1 January 2023].

Bello, C 2023 ChatGPT: AI will shape the world on a scale not seen since the iPhone revolution, says OpenAI boss. *Euronews*, 25 January. https://www.euronews.com/next/2023/01/25/chatgpt-ai-will-shape-the-world-on-a-scale-not-seen-since-the-iphone-revolution-says-opena [Last Accessed 26 January 2023].

Benzine, V 2022 A.I. Should Exclude Living Artists From Its Database,' Says One Painter Whose Works Were Used to Fuel Image Generators. *Artnet* 20 September. https://news.artnet.com/art-world/a-i-should-exclude-living-artists-from-its-database-says-one-painter-whose-works-were-used-to-fuel-image-generators-2178352 [Last Accessed 1 January 2023].

Bianchi, F, Kalluri, P, Durmus, E, Ladhak, F, Cheng, M, Nozza, D, Hashimoto, T, Jurafsky, D, Zou, J and **Caliskan, A** 2022 Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. DOI: https://doi.org/10.1145/3593013.3594095

Birhane, A, Prabhu, V U and **Kahembwe, E** 2021 Multimodal datasets: misogyny, pornography, and malignant stereotypes. DOI: https://doi.org/10.48550/arXiv.2110.01963.

Borji, A 2022 Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. DOI: https://doi.org/10.48550/arXiv.2210.00586

Bridle, J 2022 Status [Twitter]. 8 September. https://twitter.com/jamesbridle/status/1567794888103559171 [Last Accessed 1 January 2023].

Brundage, M 2018 *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. Future of Humanity Institute. DOI: https://doi.org/10.48550/arXiv.1802.07228

Buist, K 2022 The Trouble with Dall-E. *Outland*, 11 August. https://outland.art/dall-e-nfts/ [Last Accessed 1 January 2023].

Carter, R 2021 *Waveform*. http://richardacarter.com/waveform/ [Last Accessed 1 January 2023]. DOI: https://doi.org/10.33008/IJCMR.202017

Carter, R 2022 *Orbital Reveries*. http://richardacarter.com/orbital-reveries/ [Last Accessed 1 January 2023].

Chinoy, S 2019 The Racist History Behind Facial Recognition. *New York Times*, 10 July. https://www.nytimes.com/2019/07/10/opinion/facial-recognition-race.html [Last Accessed 1 January 2023].

Coldewey, D 2022 A terrifying AI-generated woman is lurking in the abyss of latent space. *TechCrunch*, September 13. https://techcrunch.com/2022/09/13/loab-ai-generated-horror/ [Last Accessed 1 January 2023].

Deng, J, Dong, W, Socher, R, Li, L, Li, K, and **Fei-Fei, L** 2009 ImageNet: A large-scale hierarchical image database. DOI: https://doi.org/10.1109/CVPR.2009.5206848

Denton, E, Hanna, A, Amironesei, R, Smart, A, and **Nicole, H** 2021 On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society*, 82. DOI: https://doi.org/10.1177/20539517211035955

Dorsen, A 2022 AI is plundering the imagination and replacing it with a slot machine. *The Bulletin of the Atomic Scientists*, 27 October. https://thebulletin.org/2022/10/ai-is-plundering-the-imagination-and-replacing-it-with-a-slot-machine/ [Last Accessed 1 January 2023].

Edwards, B 2022 Stability AI plans to let artists opt out of Stable Diffusion 3 image training. *Ars Technica*, 15 December. https://arstechnica.com/information-technology/2022/12/stability-ai-plans-to-let-artists-opt-out-of-stable-diffusion-3-image-training/ [Last Accessed 1 January 2023].

Grant, T 2014 *When the Machine Made Art: The Troubled History of Computer Art*. London: Bloomsbury.

Haveibeentrained undated. https://haveibeentrained.com/ [Last Accessed 1 January 2023].

Heikkilä, M 2022 This artist is dominating AI-generated art. And he's not happy about it. *MIT Technology Review*, 16 September. https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/ [Last Accessed 1 January 2023].

Leslie, D 2020. Understanding Bias in Facial Recognition Technologies: An Explainer. *The Alan Turing Institute*. DOI: https://doi.org/10.2139/ssrn.3705658

Liu, B et al. 2018 Beyond Narrative Description: Generating Poetry from Images by Multi-Adversarial Training. DOI: https://doi.org/10.1145/3240508.3240587

Liu, Y. et al. 2020 Generating Chinese Poetry from Images via Concrete and Abstract Information. DOI: https://doi.org/10.1109/IJCNN48605.2020.9206952

Loab.ai undated. https://loab.ai/ [Last Accessed 1 January 2023].

Offert, F 2022 *Ten Years of Image Synthesis*. https://zentralwerkstatt.org/blog/ten-years-of-image-synthesis [Last Accessed 1 January 2023].

Oppenlaender, J 2022 Prompt Engineering for Text-Based Generative Art. DOI: https://doi.org/10.48550/arXiv.2204.13988

Parikka, J 2019 Inventing Pasts and Futures: Speculative Design and Media Archaeology. In: Roberts, B, and Goodall, M (eds.) *New Media Archaeologies*. Amsterdam: Amsterdam University Press. pp. 205–232. DOI: https://doi.org/10.1017/9789048532094.010

Perrigo, B 2023 OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. *TIME*, 18 January. https://time.com/6247678/openai-chatgpt-kenya-workers/ [Last Accessed 20 January 2023].

Raieli, S 2022 Blending the power of AI with the delicacy of poetry. *Medium*, 30 June. https://towardsdatascience.com/blending-the-power-of-ai-with-the-delicacy-of-poetry-3671f82d2e1 [Last Accessed 1 January 2023].

Rettberg, J et al. 2022 Representations of machine vision technologies in artworks, games and narratives: A dataset. *Data in Brief*, 42. DOI: https://doi.org/10.1016/j.dib.2022.108319

Rettberg, J, et al. 2021 *Machine Vision in Art, Games and Narratives*. http://machine-vision.no. [Last Accessed 1 January 2023].

Roose, K 2022 An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy. *New York Times*, 2 September. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html [Last Accessed 1 January 2023].

Salvaggio, E 2022 How to Read an AI Image. *Cybernetic Forests*, 2 October. https://cyberneticforests.substack.com/p/how-to-read-an-ai-image [Last Accessed 1 January 2023].

Supercomposite 2022 Status [*Twitter*]. 6 September. https://twitter.com/supercomposite/status/1567162288087470081 [Last Accessed 1 January 2023].

Tsirikoglou, A, Eilertsen, G, and **Unger, J** 2020 A Survey of Image Synthesis Methods for Visual Machine Learning. *Computer Graphics Forum*, 396: 426–451. https://onlinelibrary.wiley.com/doi/full/10.1111/cgf.14047. DOI: https://doi.org/10.1111/cgf.14047

VanderVeen, Z. 2010 Bearing the lightning of possible storms: Foucault's experimental social criticism. *Continental Philosophy Review*, 43: 467–484. https://link.springer.com/article/10.1007/s11007-010-9160-7. DOI: https://doi.org/10.1007/s11007-010-9160-7

Williams, A, Miceli, M, and **Gebru, T** 2022 The Exploited Labour behind Artificial Intelligence. *Noema*, 13 October. https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/ [Last Accessed 1 January 2023].

Xue, Y et al. 2022 Deep image synthesis from intuitive user input: A review and perspectives. *Computational Visual Media*, 81: 3–31. DOI: https://doi.org/10.1007/s41095-021-0234-8