The Computational Approach To Aesthetics: Value Alignment And The Political Economy Of Attention In Museums And Text-To-Image Generator Stable Diffusion

Sami P. Itävuori

❧ Abstract

Whilst research into cultural value and digital technologies is nascent in art museums, neural media technologies like generative AI pose new methodological and theoretical challenges. Looking at the case of the Tate Gallery and the dataset LAION 5B used to train the text-to-image Stable Diffusion model, the article highlights the long running challenges of studying digital media from a museum perspective. Reflecting on previous uses of AI in the museum, they propose experiments in dataset research and analysis by which museums can evidence the use of their images in the training of Stable Diffusion. But these experiments also aim to develop ways in which changes in cultural value can be analysed and theorised when art collection photographs get operationalised in LAION 5B. Sketching the first steps of an epistemological analysis of image aesthetic assessment and aesthetic predictors from the perspective of museum values and aesthetics, I call for a more thorough engagement with the discourses and practices on art developed in computer sciences so that new collective and connected imaginaries of culture and advanced technology may be constructed.

Introduction

Art museums have tended to frame their understanding of AI as an add-on to existing museum activities and a tool to fulfil legacy missions like conservation, collection activation, museum learning or productivity gains. This utilitarian approach has clear benefits in supporting various aspects of museum work. But focussing too closely on this dimension disregards not only the heterogeneous systems that fall under the umbrella term 'AI' but also the economic, social and cultural forces shaping it. Whilst museums like Tate have taken small steps in this direction with projects such as Transforming Collections, a more focused discussion of images and AI, aesthetics and technology, values and automation can help address "the yawning gap" (Rutherford 60) that is said to separate the mindsets of museum practitioners and computer engineers.

Research into this area is crucial to fully understand the technological ecosystem that art institutions such as the Tate gallery participate in and to inform their public programming and curatorial practices, considering the emerging digital politics of generative AI systems. Museums also raise significant problems about the formation of aesthetic and cultural value around aesthetics and visuality as they get conceived in the case of Stable Diffusion, a popular generative AI model and digital image generation service that is widely used in commercial systems such as Midjourney or DreamStudio.

The aim of this article is to demonstrate the use of images issued from the Tate's collection in the training dataset of Stable Diffusion (SD) and briefly explain how this influences the production of images generated by this model. Once this link has been established, I will argue that divergent sets of values regarding art and its purpose emerge from the computer sciences literature in what I tentatively propose to call a *computational approach to aesthetics*. In relation to SD, this approach is not only a set of computational image-processing techniques that radically change the contextual use and nature of images harvested from the net. Instead, these generative techniques also produce and interpellate subjects as objects of scientific research and automation, as well as producers and consumers of data. Not only does SD rely on the automation of creative and cognitive tasks previously performed by humans, but its operations are predicated on prior modes of attention capture and commodification that underlie current digital platform economies (Nixon). Before undertaking this critical discussion, I will shortly survey existing approaches that have been adopted to generative AI within the museum sector in what I think are archetypal examples.

AI in the Museum

AI's areas of application in the museum are numerous but I will be focusing on projects involving digital collections of art and how AI has been used to open the collection or create new ways of searching it.

Custom AI models such as the *Digital Curator* (2022) have been built to retrieve patterns within large amounts of collection data from a consortium of Central European art museums. The browser-based platform enables users to see the statistical occurrence of objects such as "melons" or "monsters" by period, geography or artistic movement. It opens the metadata of collection images to visitors whilst also incentivizing them to discover lesser-known artefacts.

This idea of the discovery of new images is also present in other projects. The Rijksmuseum's *Art Explorer* (2024) invites users to write prompts regarding their current emotional state, their likes and dislikes into a browser-ran generative pre-trained transformer model (GPT) that then retrieves assets from the digitised collection. The interface aims to create a more intimate and affective approach to the collection, surfacing works the user wouldn't have intuitively searched for themselves.

On the other end of the spectrum in the context of Helsinki Biennale, the *Newly Formed City* (2023) project deployed AI to curate an online exhibition where artworks from the Helsinki Art Museum's collection are located on a web mapping platform of the city. The artworks are algorithmically selected, placed on a digital map and the digital images of paintings or sculptures get inserted into the panoramic digital street views of these locations. This model applies a filter to the surrounding landscape, like augmented reality apps. The filter transposes the artwork's formal qualities such as colour, texture, materials, shapes onto the digital landscape, providing a new experience of the city to local inhabitants and visitors alike.

Many more examples have been recorded in recent literature on the topic of museums and AI more generally. In 2021 Soufian Audry highlighted the emergence of AI as a popular topic for museum exhibitions, surveying eight international exhibitions on the topic (4).

Similarly, the edited volume *AI in Museums* (ed. Thiel and Bernhardt) present the various applications of AI in all areas of museum work from collection management to education via marketing and curation. Hufschmidt surveys a hundred and twenty-two such projects taking place between 2014 and 2019, with most of them focusing on enhancing visitor experience of the collection with audio guides and collection search-tools (133). Most recently, the 2025 MuseumNext's *MuseumAI Summit* brought together international museum professionals and creative technologists with a focus on collection activation and visitor-data analysis using AI.

Museums are actively adopting AI-powered software to analyse or 'activate' the vast amount of data they hold about objects in their collections. But it is very much of 'adoption' that I am talking about here, insofar the techniques of artificial intelligence are adopted from outside the museum by either using off-the shelf models or commissioning creative technologists to do it for them. There is nothing

inherently wrong about these approaches given the technical complexity of these models and the significant skills and investment custom models require. But this nevertheless isolates the museum and the development of AI products from each other as separate fields of life, activity and reflection, with little to no common ground for dialogue or interrogation.

This separation is not without consequences and contributes to a "yawning gap" between the concepts, mentalities and practices of museums and of developers behind AI systems (Rutherford 60). It reinforces the 'black box' narrative that hinders a head-on engagement with the 'AI tech stack' (Ivanova et al.) as too complex or too big for scrutiny by researchers outside computer sciences (Bunz 26; Gogalth 175). Whilst issues of bias, privacy or copyright are already being discussed in the sector, art museums are largely lacking means to stir a critical reflection on the inherently social and economic dimensions of AI technologies, their impact on human lives and the role of the museum collection in an era of neural media. Considering AI in the singular mystifies the variety of techniques that are deployed to analyse and synthesise large amounts of data, and the values that guide these deployments. It leads the conversation away from the real problem: the human use of these technologies with and on other humans. The digital politics of museum collections in an information society, the process of defining the values that guide their existence and societal role, thus need to be revised considering emerging AI powered neural medias (Fuchsgruber; Allado-McDowell)

This idea of the museum having societal agency builds on the contemporary articulation of its role as not only sites of collection preservation and exhibition, but also spaces of experience centring the visitor, their needs and agency in what has been called the post-museum (Hooper-Greenhill 22). With roots in the nineteen-nineties new museology, this re-centring of the visitor and the civic role of the museum in the UK had also been pushed since the two-thousands by the focus on culture's "use-value" in securing government funding (McPherson 46). This reformatted the museum as a site of pedagogy and entertainment, to both address the growing competition of new media and experience economies for public attention, as well as produce measurable impact metrics to justify public funding of these institutions (Scott, Dodd and Sandell 9). Whilst this policy orientation has pushed a new industry of quantitative research about public impact and outreach, the museum object remains conceptualised as holding an 'intrinsic' value that ties personal experience to collective meaning making. In this definition, the artefact and museum expertise (organisational, pedagogic and curatorial) mediate the representation of a social group to a symbolic world linking the past to the present as well as a potential future, endowing heritage institutions with a unique societal role (Crossick and Kaszynska 16).

Whilst this haptic dimension of the museum's symbolic function is a constant in the justification of museum collections and investment in preservation work, this intrinsic value of the object has been complicated by digital technologies that create distance between the audience, the space and the object, but also new

distributed modes of communication about images, stories and experiences of artworks. The meanings and contexts of artworks have been further fragmented in digital networks for instance. These networks multiply the sites and voices that mediate the reception and discussion of artefacts, and trouble the institution's curatorial authority, which often relies on one-way broadcasting modes of online communication (Styles; Zouli). Online media landscapes complicate not only the measurement, but the very conception of cultural value, as the parameters of art and its images' experience change (Dewdney and Walsh 15). A trend that only seems to be accentuated by the creation of machines capable of identifying, evaluating and recreating images of art and whose values seem to conflict with values associated with artistic authenticity, creative labour or the disinterestedness of aesthetic experiences. There seems to be an inherent problem with the 'alignment' of values between institutional perceptions of art in museum collections and emerging generative AI, which build on previous tensions from preceding digital medias like television or the internet.[1]

So, what is the transformation of cultural value that has taken place with the advent of systems that can produce images of art with natural language text prompts? And what does this say about the emerging relation of art museums to these models?

To answer this question, I will now evidence the link between Tate gallery and the text-to-image generative AI model Stable Diffusion.

Diffused Images

i. the digital photograph of artworks

The Tate gallery is a national museum in the UK that consists of four geographical sites across London, St. Ives and Liverpool. To use language from its previous media strategies in the early parts of the two-thousands, Tate's website was conceived as the "fifth site" of Tate with its own programme and dedicated visitor resources (both curatorial experiment and "brochure ware") (Rellie). The history of this "fifth site" can be traced back to the British Art Information Project (BAIP) of the late 1990s, which promoted the large-scale digitization of collections and archives across national portfolio institutions in the UK. Whilst digital photography of collections started in the early nineties, the launch of the Tate's website in 1997 is directly tied to this digitization project in the build-up to the opening of the new Tate Britain wing as part of the museum's Millenium Project.

The current iteration of the collection website *Art and Artists* displays more than seventy-seven thousand collection photographs ranging from film stills to paintings via artist sketches and installations. Most photographs are available under creative commons licenses or can be licensed for a fee from Tate. Guiding the release of these images online was the idea of supporting access to the collection, regardless of geographical and temporal boundaries. It supported the fundamental targets of this national museums' mission statement to promote the appreciation and

understanding of British and international art to the public in addition to the preservation of collections in digital form. Adjacent to these aims, the net was also conceived as a site of free information circulation, as well as an expanded marketplace where virtual visits would translate into 'real' footfall and income in the gallery (ticketing, catering, gift shop).
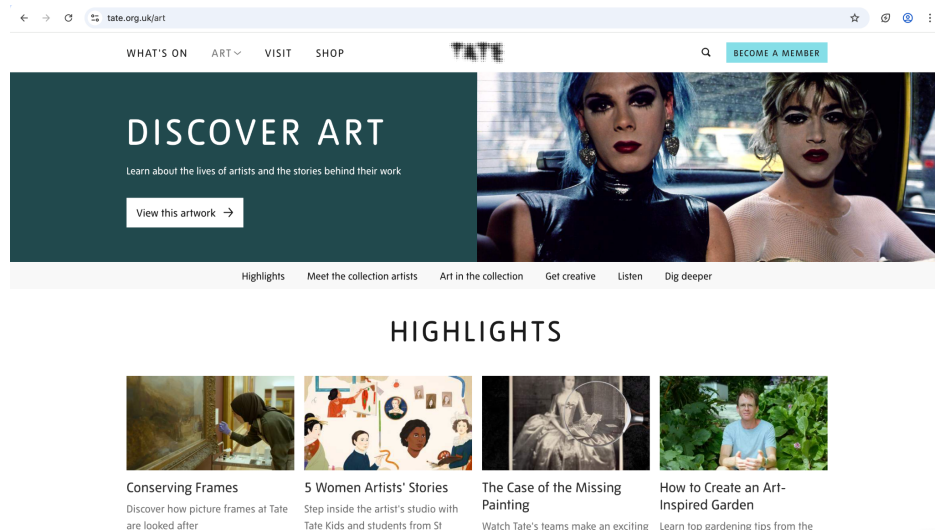

Figure 1: Screenshot of the Art and Artists Section of the Tate Website. https://www.tate.org.uk/art (accessed 5. June 2025).

Digital media then becomes a ground where disparate, overlapping values get negotiated: tools to support the cultural and civic mission of the museum to promote its art collection as valuable in and for itself. Tools supporting governmental goals of societal regeneration and education. But also, tools to support the emerging entrepreneurial business model of the museum following public funding cuts (Hughes 9). The production of digital photographs of artworks in museums is thus guided and animated by these divergent and concurrent values that co-exist when being operated internally in organisational databases and circulated on the public facing website.

The online circulation of the artwork's digital image leads to a change in its nature. Whilst the physical artwork remains mediated and 'framed' by the institutional discourse and values, the digital photograph of the artwork once online gets appropriated, reused and maybe misused in a multitude of ways by online users. The circulation of digital images means that any image posted online undergoes endless copying, compression, editing and pasting that impact the image technically (degraded resolution, dimensions, watermarks, captions), and culturally by decontextualising the image (Steyerl, "In Defence of the Poor Image"). Once online, the museum relinquishes a degree of control over its reception and uses, including its reproduction, derivations or commodification. This network of online circulation is the ground on which the museum meets new AI systems such as Stable Diffusion, which I will briefly present now.

ii. the images of Stable Diffusion

Stable Diffusion (SD) is a popular generative AI system developed by the Ludwig-Maximilian University of Munich and the British private company Stability AI. The system can generate images from natural language prompts. SD operates on a diffusion model (DM), which is a process of deep learning (Rombach et al). In training a DM, a set of algorithms called a neural network is iteratively improving its capacity to remember the content of 'images' and to reconstitute them from statistical noise. How does it do this? The DM relies on a process of noising – where the data in input images is increasingly deteriorated by inserting Gaussian noise. The aim of the model is to then denoise the degraded image by following successive steps of reconstructing the target image (or 're-membering', putting parts or pixels back into their place). The noisy starting point has some structured clusters of pixels remaining in it, which the model builds on to draw the outlines of its target image.

How is this prediction process guided? DM utilises a pre-trained machine vision model called the Contrastive Language-Image Pre-training (CLIP), which ties textual descriptors to image data within a latent space (Radford et al). A latent space is a high-dimensional statistical space where image data and text data are converted into machine-readable numerical tokens. These tokens act almost like coordinates on a 3D map and each token's position is determined by the statistical frequency of their co-occurrence in the training data. This is essentially the model's 'attention' to the context-specificity of certain words and figures, and how it can differentiate the 'apple' in a tree from 'Apple' computers. This is also how the model can be steered to produce new images that may not exist in its dataset by writing text-prompts. The prompts connect different areas of the latent space and enable a hybridisation of the data. The model aims to guess how these images would look like and rely on the successive feedback of humans but also automated models to either validate or reject its predictions and to re-adjust its process. The illustration in Figure 2 aims to illustrate the prediction process of SD, denoising a seed image in 4 stages for the prompt "photorealist image of an apple-computer (1 steps, 4 steps, 5 steps, 15 steps).
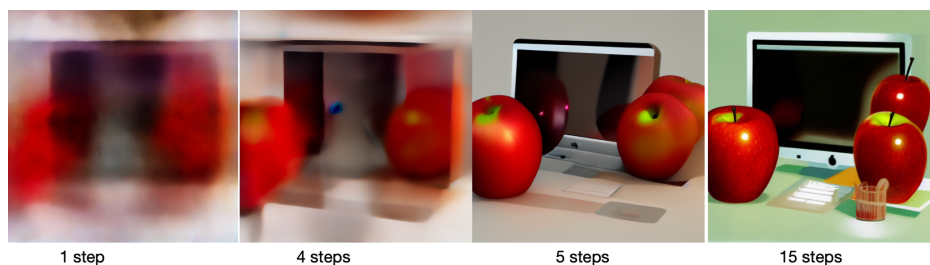

Figure 2: four denoising steps on Stable Diffusion (v. 2023) ran on ComfyUI. https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0/tree/main

In this broad summary, I hope it is clear that generative models like SD rely on the association of natural language descriptors to image data. Also essential is the

compression of this data into a new tokenised form, distributed in a multidimensional vector space according to the statistical frequency of their occurrence. By using language and mathematical abstraction, a semiotic-visual syntax and vocabulary is developed. The model is seemingly able to 'know' what an apple or a computer looks like and thus be able to 'predict' what an Apple computer or image an 'apple-computer' must look like, based on its 'memory' and attention to context in the latent space.

It is inevitable that the popularity of this diffusion model will be supplanted by another model in the years to come, especially given the speed at which research is moving in the current AI arms-race (Aschenbrenner). I would still argue that the fundamental representational logic of generative image-making is unlikely to change. This logic relies on the collecting, modelling and articulation of images with language and mathematics to make them machine-readable. Whilst in contemporary art theory semiotics had been mostly evacuated from the visual field, particularly with the 'autonomy' thesis of the artwork and the challenged 'indexicality' of photographic images, the concepts behind diffusion models take a diametrically opposite approach. In diffusion models, images are framed as solid representations of a homogeneous reality that is describable by language and statistics. This correspondence of images to reality in DM raises a series of broader epistemological questions about images of art, their treatment in generative AI research and the underlying aesthetic culture of this technology. On this basis it is enlightening to look at the ways images are pre-processed for models like Stable Diffusion in their visual memory bank, namely the dataset LAION 5B.

iii. Images of LAION

Here the online circulation of collection images ties into the pipeline of machine learning training of SD. The training process requires large amounts of image-text data, which in the case of SD were harvested from the internet using a bot called a crawler. Whilst the initial crawling process of downloading large swaths of the internet was done by a not-for-profit organisation called Common Crawl, it was another not-for-profit, LAION which with the support of Stability.AI and the LMU curated and compiled the LAION dataset (Schuhmann et al). Currently LAION-5B is a five billion strong dataset with images and text harvested from the internet. It contains both the 'best' and the 'worst' of the internet, from amateur websites to stock photos, photographs posted on Flickr or digitised artworks. In my research at Tate, I chose to work on diffusion models like SD because the training dataset LAION is available open-access and thus searchable. Until recently LAION-5B was available for free download on Hugging Face but in 2024 significant amounts of harmful, abusive and illegal content was discovered, forcing its removal from circulation until a full audit is completed (Thiel 7). The sheer amount of data contained in LAION means that having the images checked by humans was considered impractical, too lengthy and too expensive. For these reasons the reviewing process was automated with an object recognition system. The system was tasked to rate the 'safety' of images depending on their likelihood to have

harmful content or poor image quality. LAION's five billion images had been algorithmically reviewed. The algorithm had failed on several occasions.

A smaller subsection of LAION considered to be safer and of a 'higher aesthetic quality' is still online and available to download. This LAION-Aesthetic subset contains eight-million lines of data. Each item includes certain essential information including the source URL of the images and text.

I coded a simple search tool to identify assets issued from the Tate website. Browser-run search-tools had previously been available for LAION-5B until it had been taken offline. I settled on the use of Datasette, an open-source software that enables the search of datasets using the SQLite relational database model. The software was coded and ran using the cloud-based development environment Github Codespace. I used a free plan and ran the program from my laptop using a 2-core, 8GB RAM, 32GB virtual machine.

Because LAION-5B and LAION-Aesthetic are presented as relational datasets made of columns holding asset metadata, I had the possibility of filtering LAION-Aesthetic by search terms. As each image-text pair was indexed with their source URL, I was able to filter the database for the domain "tate.org.uk" and received 354 results back. The scraped images ranged from works by J. M. William Turner (thirty-two in total), a photograph of the sculpture *Winter Bears* (1998) by Jeff Koons or illustrations by Beatrix Potter. Nine women artists were represented out of a total of 125.

Figure 3: Searching for URLs matching "tate.org.uk" on LAION Aesthetic using Datasette.

Fourteen works were from the 1700s,147 from the 1800s, 132 works from the 1900s. The rest of the images did not contain a date in their text data. All images, to the exception of *Winter Bears,* were photographs of two-dimensional paintings. These figures largely reflect the make-up of Tate's collections. For instance, the Tate holds 37,000 works and sketches by William Turner, the majority of which has been digitised.

This selection of images illustrates a data bias towards representing works from the art historical canon. The type of images present in databases like LAION are an essential part of the conversation on the biases of AI systems, which has been the subject of significant attention in critical AI scholarship and computer sciences (Ferrara 2). This bias is also recognised by the developers of SD: "deep learning modules tend to reproduce or exacerbate biases that are already present in the data" (Rombach et al. 9). But saying that the bias lies "already" in the data seems to ignore biases that occur when programmers process, mediate and operationalise the collected data (Offert and Bell 1133). By this I mean that LAION is not just made up of raw data collected from the wild. Instead, a series of human decisions based on cultural and technical rationales determine what data gets used, how and why. For this reason, models like SD don't generate new media just out of raw data but are deeply informed by decisions underlying the collection of data, human interpretation of this data and the aims they want to achieve with it. The bias is already in the human process of capturing the world as information in what could be called the 'capta' (Drucker 2).

Figure 4: Image of Apples in LAION Aesthetic. Described as 'apples'. Aesthetic Score 818764114379828. Source: Bird Feeder Expert website, https://birdfeederexpert.com/wp-content/uploads/2020/12/orchard-1872997_640.jpg

Looking closer at other categories of the dataset the column "aesthetic" stands out. "Aesthetic" denotes the aesthetic score attributed to each image on a sliding scale of zero to ten to measure its aesthetic quality. This quality score is a prediction of the image's appeal to a human viewer and the scores categorise images in LAION from poor to good quality. The images are distributed in 'buckets', that is groupings of images by score. The lowest buckets are poor resolution images with watermarks for example and those which supposedly contain potentially harmful content, whilst the high-quality images are in the buckets 8 to 10. Thus, images are not only described in terms of their content: apples. They are also rated:

this image of apple scores 8.18764114379828. This is very interesting at two levels: how are images evaluated for their aesthetic quality? And why are they evaluated? It is to these questions I turn attention to in the next section.

Image aesthetic assessment: virtual viewers and platform capitalism

Just as they are analysed for perceived harmful content (their safety score), images in LAION are allocated an aesthetic score automatically using a modified version of the CLIP model called an aesthetic predictor (Schuhmann & Beaumont).

Two datasets were used in CLIP aesthetic predictor's pre-training, namely the Simulacra-Aesthetic Captioning (SAC) and the Aesthetic Visual Dataset (AVA). Both datasets are considered benchmarks in machine vision research. Both SAC and AVA contain photographs scored by humans either on online photo competition platforms or by research participants in academic studies. SAC contains ratings for 230 000 AI-generated images (Pressman), whilst AVA contains 250 000 images with ratings and comments collected from the photo-challenge platform DP.Challenge (Murray et al.). Both SAC and AVA were used in training the aesthetic predictor for LAION and thus inform the production of SD's memory and its parameters for evaluating appeal. But how is this appeal defined and determined? Especially since appeal appears at first as a subjective phenomenon.

Underlying the automation of aesthetic rating, lies the scientific effort to theorise and measure the impact of images on humans or elucidate the qualities that make an image appealing. The field of Image Aesthetic Assessment (IAA) takes up this question, with research being undertaken in neurosciences, cognitive psychology, computing and marketing research (Bodini 5). The aim of IAA is to determine what makes an image beautiful and to produce experimental apparatuses, including computational simulations, to support these theories. Surveys of the literature show significant previous efforts to develop an automated assessor based on the formal analysis of artworks. This means hand-crafting arbitrary lists of positive visual qualities, such as composition, subject, rule of thirds, colour combinations, contrast and more (Deng, Loy and Tang, 6). This approach however has lost in popularity since the appearance of deep neural networks that utilise large amounts of data and bypass the need for hand-selected features. The idea that there are a-priori formal qualities underlying the aesthetic appeal of images could be called an objectivist approach to the study of aesthetics (Bodini 4), because it centres the object's qualities as the source of subjective aesthetic experiences. In contrast, current deep learning models aim to mimic the behaviour of human test-groups without prior knowledge of formal qualities that make a visual object appealing. In this subjectivist approach, the impact of an image on its observer takes precedence over formal qualities. In general, this approach aims to measure the impact of individual images on a scale from zero to ten on thousands of human test subjects to produce average scores on a large body of images (Folgerø 19). A deep neural network is then trained to start recognizing patterns of pixels that tend

to be associated with high human scores. These minute pixel patterns that are invisible to the human eye are the building blocks from which the machine perceives appeal, a process that seems completely opposite to human ways of perceiving and receiving images as totalities, rather than minute details. In this paradigm, appeal is not a tangible objective quality but a statistical trend that 'emerges' from the aggregate of human ratings collected 'in the wild'. Echoing Chris Anderson's controversial 2008 claim that big data marks the end of theory in scientific research, current IAA paradigms assert that with enough data, numbers can not only explain but also make aesthetic judgements.

CLIP Aesthetic is built on this subjectivist approach because it has taken most of its scores from online photography websites. One of them, DP.Challenge, is an amateur photography competition run by the Digital Photography Review since 2002. Users are invited to organise thematic competitions and to upload images, which are then anonymously scored and commented. This is the data used for CLIP, but other researchers have proposed to use Reddit's r/photography discussion thread (Nieto et al) or Flickr comments (Soydaner et al) as alternative data sources to train IAA deep learning models. The scores are averaged, but also sometimes require correcting as simple averages tend to neglect sentiment polarity. Polarity, the presence of both very high and low ratings simultaneously, defines an image as 'divisive' because the consensus on its score is considered less reliable. These polarizing images are often removed from the training set because images that have wider appeal are considered 'truer' examples of aesthetic attractiveness since they gather unanimous agreement amongst scorers (Park and Zhang). In a sense CLIP aesthetic predictor is a simulation of user behaviour on platforms like DP.Challenge (as well as the university student test group of AVA, which should be discussed in a separate paper). The aesthetic predictor's aim is to predict reliably the appeal of images, and thus the middle of the road, or 'mean' aesthetic gets prioritised over images that may be divisive or appear unconventional to users. Underlying these aims is to make the model as popular and appealing as possible to a wide user- and consumer-base.

This notion of unreliability is important to discuss the differing value systems that guide SD and museums when they utilise digital images of art. In the training pipeline, the attribution of aesthetic scores dispels the idea of text-image data as something "already" in the data. It is a capta in the sense that the choices underlying the selection of the data, the source of the data and the truth-value assigned to this data are processes of capturing and socially interpreting information within an epistemic and technical framework. In this situation, the aim to produce appealing images with SD means that a way of formalising aesthetic appeal becomes a technical pre-requisite for the synthesis of new images. Images from museum collections are inputs for machine learning and the cognitive processes of human viewers of art also become conceived as inputs. The formalization of appeal requires its definition and in the case of CLIP-aesthetic, appeal is defined as the statistical frequency of pixel patterns unconsciously liked by online photo communities. The user ratings from DP.Challenge then constitute a

'ground truth' of aesthetic appeal (Sluis) in the development of image generators like SD. These ratings are representative of a generalizable human cognitive response to images and can be inferred to make guesses about the appeal of future images. This process is automated in CLIP and finally defines the aesthetics and visual look of images produced by SD. Generative AI then not only translates a further datafication and commodification of images of art, but a datafication of photo-competition participants' cognitive labour when they produce scores and feedback about photographs. In this sense, these systems do not only treat images from national art collections as a means to an end, but they also objectify human interactions with these images, they objectify aesthetics.

Analysed from a visual cultures or aesthetic theory perspective, this approach seems to have several problems. The inductive nature of the reasoning behind the operationalization of this subjectivist approach is problematic because the ratings of amateur hobbyist photographers from North America are confounded with a universalizable notion of aesthetic taste (Sluis and Palmer). This generalization highlights a Western photographic unconsciousness in generative AI but also points to the strong geographic and cultural contingency of the aesthetics promoted by DP.Challenge. This also applies to CLIP aesthetic as it was trained on the same data. It could be said then, that LAION-5B is organised by an automated 'virtual viewer' of these images, a sort of 'virtual connoisseur' with median aesthetic taste, engineered by a mixture of cognitive psychology, neuroaesthetics, statistics and computing. The aesthetic predictor is a form of automata, performing human-like labour of indexing and tagging but its logic seems alien to the affective charge of human reception and interpretation of images. It is built on data about human behaviour, but the predictor's behaviour is very different from the subjective human experience of the world, which seems to change at every iteration, particularly in the case of art. Art is characterised by the difficulty to agree on its definition, but also by a polyvocality in the experience of individual artworks. To just quote a historical example used by Helliwell to discuss value alignment in AGI, the work of Vincent Van Gogh, who was one of the early artists to be used to showcase the power of style-transfer (Gatys et al 5), was derided during his lifetime. His work got positive recognition only decades after his death and points to the fact that the appreciation of artistic styles changes over time and is not a fixed quality, which can be captured at one point and reproduced indefinitely. Theoretically this means then that SD's logic would prohibit the spontaneous emergence of new visual aesthetics that do not conform to existing tastes and preferences. The algorithms then appear as deeply conservative. A reality that stands at odds with the recurrent discourse of progress, democracy and futurity invoked by the developers of these systems.

Conclusion: From Images to Attention Economy

But why is generative image-making the object of so much economic investment, artistic controversy and popular mass adoption? I argue that the techniques behind image generation actually build upon and reinforce the commodification of the

online space. It frames users as customers and any data as resources to be extracted and monetised. The subject being automated in generative AI's virtual viewer reflects this atomised subject of digital platform economies. Online users are being atomised because the platforms on which they build their online existence aim to increasingly isolate them from each other, whilst also extracting as much capital from them in the form of service income or the data they produce (Bridle, 91). This atomisation is further reinforced by these platforms' reliance, from social media to generative platform, on the commodification, capture and retention of the users' time and attention. This race to capture attention lies behind early investment in IAA research so companies could better understand how online consumers act. Internet users are simultaneously interpellated[2] as consumers by algorithmic recommendations within a wider digital marketplace for the provision of goods and services (Terranova 2; Hentschel, Kobs and Hotho 2; Baeza-Yates and Fayyad 132). The political economy of attention in communicative media discussed by Nixon ties the epistemology and techniques of aesthetic appeal in generative AI. It also ties to the wider exploitation of data produced by image-makers, artists, museums and online commentators worldwide.

This process of attention capture, extraction and image generation perpetuate data colonialism's framing of digital networked images as "an 'open' resource for extraction that is somehow 'just there' for capital" (Couldry and Mejias, 337). Previous logics of platform companies such as search engines and social media extend the creation of new consumer needs in the form of mercantile image-generation services online. By extracting digital images, companies selling generative AI services have effectively privatised the internet commons. This lies at the source of controversies with artist lawsuits, scriptwriter strikes and cross-sector concern about the future of creative economies. This process of data pillaging, data colonization and privatization runs parallel and within the scientific project to measure, define and quantify human psychological and cognitive processes to predict the appeal of images, messages and information, both analogue and now synthetic. These techniques of observation, and now generation, continue to inform a symbiotic relationship between emerging modes of visual culture, scientific study of human cognition and emerging modes of economic exploitation (Crary).

This subject atomisation in the digital sphere and reality driven by the aesthetic predictor, the visuality of SD or the marketplace of consumer-oriented platforms stands at odds with the cultural values that are usually associated with artistic heritage and its digital images. Leaving aside the ways in which generative AI models decontextualise all images in their datasets to exploit their representativeness and reduce them to textual descriptors and aesthetic scores, the political economy of these models tends to disenfranchise artists and break the symbolic function of images as a site where meaning, identity and histories are collectively negotiated, preserved or relinquished. As pointed out by Steyerl, the formation of a common-sense of aesthetics relies on the messy, asynchronous and sometimes unresolved reception of images, whose attraction may endure even if their appeal is polarizing (*Medium Hot*, 51). This raises questions about the

possibility of alignment of museum missions with the emerging visuality of generative AI.

Despite their creators' aspiration to "democratize high-resolution image synthesis" (Rombach et al. 1), the inherent political economy of the aesthetics of systems like SD appear more to alienate than to strengthen social bonds and promote creativity. LAION's virtual viewer is not fully the same viewer as the human in the gallery space or the museum website. The instability and free-play associated with aesthetics gets reduced to rational choice-making subjects modelled according to contemporary market logics in aesthetic predictors. Not that market logic is exclusive to the machine, it also animates the values behind museum entrepreneurialism, although these values are constantly negotiated and problematised within the linear progressive values of the institution issued from representational artistic modernity (Dewdney, 5).

By standing at the crossroads of new forms of economic exploitation and emerging forms of human image making, generative AI problematises what it means to look at images, where we look at them and what infrastructures facilitate these modes and techniques of vision. It also raises questions regarding art, images and private property. The process of privatizing data issued from the internet commons reframes all data as a resource to extract and own, thus determining who gets to monetise it, when and how (Bailkin, 14). A similar paradigm of property ownership is characteristic of the way in which museum collections operate, the exhibition being a format that conditions what works can be seen, how and when. But museum websites have gradually disrupted this property paradigm. For instance they have aspired to open the collection's stores and promote the idea of artwork stewardship: namely that artworks in collections are not mere possessions but rather common goods that need to be managed for the good of *all* (Cheng-Davies, 290). Common goods understood as that which benefits the community of users, be it in the promotion of social cohesion, provision of education, improvement of mental health or some other projected value of the artwork. Whilst the common good can be seen to animate museum activities when they release digitised collection data online and partake in public programming, it is less evident in the way that commercial generative AI platforms use images issued from national collections. The opacity of platforms like Midjourney or Dall-E that hides behind convenient user-interfaces, reinforces barriers to a wider understanding of how these systems work and the economic processes that make them possible. This then poses two problems for the humanities and museum institutions: how can cultural institutions reassert a common capacity of anyone to understand and tinker with these systems? What mechanisms or imaginaries need to be formulated, and by whom, to redistribute the benefits of these generative technologies for a common good: creatively, societally, financially?

The task at hand is then to develop strategies, curatorially, organisationally or infrastructurally that promote the reappropriation of data heritage, digital commons and a social ownership of the means of prediction. As a site of heritage, with

buildings, expertise, objects and archives, the museum has the affordance to bring strangers together and maybe turn them into neighbours and community-members (Balshaw). This coming together is essential, not only to share grievances but also a common capacity to deliberate what the common good looks like in a specific situation, place and time. Even if the museum is defined by conflicting values including strong market forces, it also values criticality as a mode of culturally engaging with their histories and collections. This means there are affordances in the current value-system and infrastructure of the museum to engage in this conversation about re-commoning data heritage. As argued at the beginning of this article, the starting point needs to be an understanding of the technology from the perspective of those affected by it: institutions, humans, communities and cultures. As Katz writes: "Explanation of rules is a prerequisite for the democratic control of rules". (22) What this democracy may look like in the museum remains to be imagined and opens a line of research into the techno-aesthetics of generative AI models, not only to inform new museum activities but maybe assert the right and necessity of cultural workers to have a say in the ideas and applications of these fast-moving techniques. It is thus not only the rules guiding the algorithms that need to be explained but also the rules of the emerging political economy of corporations and digital platforms powered and guided by AI that need to be elucidated within visual culture.

Works Cited

– Allado-McDowell, K. "Am I Slop? Am I Agentic? Am I Earth: Identity in the Age of Neural Media." *Long Now*, 19 Feb, 2025, longnow.org/ideas/identity-neural-media-ai/. Accessed 12 June 2025

– Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete". Wired, 23 Jun, 2008. https://www.wired.com/2008/06/pb-theory/. Accessed 15 July 2025.

– Althusser, Louis. "Ideology and Ideological State Apparatuses (Notes Towards an Investigation)". Lenin and Philosophy and Other Essays. Monthly Review Press. 1971. Pp.127-188

– Aschenbrenner, Leopold. *Situational Awareness*. 2024, situational-awareness.ai/. Accessed 12 June 2025.

– Audry, Sofian. *Art in the Age of Machine Learning*. MIT Press, 2021, direct.mit.edu/books/monograph/5241/Art-in-the-Age-of-Machine-Learning. Accessed 12 June 2025

– Bailkin, Jordanna. *The Culture of Property: The Crisis of Liberalism in Modern Britain.* 2004. University of Chicago Press

– Balshaw, Marina. *Gathering of Strangers: Why Museums Matter*. Tate Publishing, 2024.

– Bodini, Matteo. "Will the Machine Like Your Image? Automatic Assessment of Beauty in Images with Machine Learning Techniques." *Inventions*, vol. 4, no. 3, 2019, p. 34, doi.org/10.3390/inventions4030034. Accessed 12 June 2025

– Bunz, Mercedes. "The Role of Culture in the Intelligence of AI." *AI in Museums: Reflections, Perspective and Interpretations*, edited by Sonja Thiel and Jonathan Bernhardt, Transcript Verlag, 2023. library.oapen.org/bitstream/id/3ecbc4ec-2dac-4e05-881a-10414c20f7f2/9783839467107.pdf. Accessed 12 June 2025

– Cheng-Davies, T. S. L. "A Work of Art is Not a Barrel of Pork: The Relationship Between Private Property Rights, Moral Rights Doctrine and The Preservation of Cultural Heritage". *Intellectual Property Quarterly*, 2016(3),

278-294 https://research-information.bris.ac.uk/ws/portalfiles/portal/86360756/article_IPQ_27_april_2016.pdf Accessed 12 June 2025

– Crary, Jonathan. *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. MIT Press, 1992.

– Crossick, Geoffrey, and Patrycja Kaszynska. *Understanding the Value of Arts & Culture: The AHRC Cultural Value Project*. Arts and Humanities Research Council, 2016, ukri.org/wp-content/uploads/2021/11/AHRC-291121-UnderstandingTheValueOfArts-CulturalValueProjectReport.pdf. Accessed 12 June 2025

– Deng, Y., et al. "Image Aesthetic Assessment: An Experimental Survey." *IEEE Signal Processing Magazine*, vol. 34, no. 4, July 2017, pp. 80-106, doi:10.1109/MSP.2017.2696576. https://arxiv.org/pdf/1610.00838 Accessed 12 June 2025

– Dewdney, Andrew. "Art Museum Knowledge and the Crisis of Representation." *Representing Art Education: On the Representation of Pedagogical Work in the Art Field*, edited by Carmen Mörsch et al., Zaglossus, 2017.

– Dewdney, Andrew, and Victoria Walsh. "Temporal Conflicts and the Purification of Hybrids in the 21st-Century Art Museum: Tate, a Case in Point." *Stedelijk Studies*, no. 5, Fall 2017.

– Drucker, Johanna. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly*, vol. 5, no. 1, 2011.

– Ferrara, Emilio. "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies." *Sci*, vol. 6, no. 1, 2024, p. 3, doi.org/10.3390/sci6010003. Accessed 12 June 2025

– Folgerø, Per Olav. "Introduction: Representative Foci in Neuroaesthetics—Subjectivist, Objectivist, and Interactionist Perspectives." *Neuroaesthetics: A Methods-Based Approach*, edited by Tudor Balinisteanu and Kerry Priest,

Palgrave Macmillan, 2024, doi.org/10.1007/978-3-031-42323-9_1. Accessed 12 June 2025

– Fuchsgruber, Lukas. "Network Culture and Online Collections, Theory for the Politics of Digital Museums." *Nullmuseum*, 26 July 2024, nullmuseum.hypotheses.org/884. Accessed 12 June 2025

– Gatys, Leon A., et al. "A Neural Algorithm of Artistic Style." *Computer Science*, vol. 11, 2015, pp. 510-519, arxiv.org/abs/1508.06576. Accessed 12 June 2025

– Golgath, Tabea. "The Funding Program LINK—AI and Culture Five Lessons Learned after Five Years." *AI in Museums: Reflections, Perspective and Interpretations*, edited by Sonja Thiel and Jonathan Bernhardt, Transcript Verlag, 2023.

– Helliwell, Alice C. "Aesthetic Value and the AI Alignment Problem." *Philosophy & Technology*, vol. 37, no. 129, 2024, nul.repository.guildhe.ac.uk/id/eprint/2226/. Accessed 10 June 2025.

– Hooper-Greenhill, Eilean. *Museums and their Visitors*. Routledge, 1994.

– Hufschmidt, Isabel. "Troubleshoot? A Global Mapping of AI in Museums." *AI in Museums: Reflections, Perspective and Interpretations*, edited by Sonja Thiel and Jonathan Bernhardt, Transcript Verlag, 2023, p. 133, library.oapen.org/bitstream/id/3ecbc4ec-2dac-4e05-881a-10414c20f7f2/9783839467107.pdf. Accessed 12 June 2025

– Hughes, Lorna M. "The Value, Use and Impact of Digital Collections." *Evaluating and Measuring the Value, Use and Impact of Digital Collections*, edited by Lorna M. Hughes, Cambridge University Press, 2012.

– Ivanova, Victoria, et al. "Art x Public AI." *Future Arts Ecosystem*, vol. 4, Serpentine R&D Platform, 2024, reader.futureartecosystems.org/briefing/fae4/preface. Accessed 12 June 2025.

– Ludwig Maximilian University of Munich. "Revolutionizing Image Generation by AI:

Turning Text into Images." *LMU Munich*, 2022, lmu.de/en/newsroom/news-overview/news/revolutionizing-image-generation-by-ai-turning-text-into-images.html. Accessed 10 June 2025.

– McPhearson, Gayle. "Public Memories and Private Tastes: The Shifting Definitions of Museums and Their Visitors in the UK." *Museum Management and Curatorship*, vol. 21, no. 1, 2006, pp. 44-57, doi.org/10.1080/09647770600602101 Accessed 12 June 2025

– Murray, Naila, et al. "AVA: A Large-Scale Database for Aesthetic Visual Analysis." *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, doi.org/10.1109/CVPR.2012.6247954 Accessed 12 June 2025

– MuseumsNext. "MuseumAI Summit." *MuseumsNext*, 26-27 Mar. 2025, museumnext.com/events/museum-ai-summit/. Accessed 9 April 2025.

– Nieto, Daniel Vera, et al. "Understanding Aesthetics with Language: A Photo Critique Dataset for Aesthetic Assessment." *Thirty-Sixth Conference on Neural Information Processing Systems*, 2022, arxiv.org/abs/2206.08614 Accessed 12 June 2025

– Nixon, Brice. "COMPASS| Critical Communication Policy Research and the Attention Economy: From Digital Labor Theory to Digital Class Struggle." *International Journal of Communication*, vol. 11, Nov. 2017, p. 13, ijoc.org/index.php/ijoc/article/view/7005. Accessed 15 June 2025

– Offert, Fabian, and Peter Bell. "Perceptual Bias and Technical Metapictures: Critical Machine Vision as a Humanities Challenge." *AI & Society*, vol. 36, 2021, pp. 1133-1144, doi.org/10.1007/s00146-020-01058-z. Accessed 25 July 2025

– Park, Tae-Suh, and Byoung-Tak Zhang. "Consensus Analysis and Modeling of Visual Aesthetic Perception." *IEEE Transactions on Affective Computing*, vol. 6, no. 3, July-Sept. 2015, pp. 272-285, doi.org/10.1109/TAFFC.2015.2400151. Accessed 25 July 2025

– Pilke, Lukas. "Digital Curator." *Digital Curator*, 2022, digitalcurator.art/aboutproject. Accessed 9 Apr. 2025.

– Pressman, J. D. "Simulacra Aesthetic Captions." *GitHub*, 2022, github.com/JD-P/simulacra-aesthetic-captions. Accessed 10 June 2025.

– Radford, Alec, et al. "Learning Transferable Visual Models from Natural Language Supervision." *International Conference on Machine Learning*, PMLR, 2021, pp. 8748-8763, arxiv.org/abs/2103.00020. Accessed 12 June 2025

– Rellie, Jemima. "One Site Fits All: Balancing Priorities at Tate Online." *Museums and the Web*, 2004, archimuse.com/mw2004/papers/rellie/rellie.html. Accessed 14 February 2025.

– Rijksmuseum. "Art Explorer." *Rijksmuseum*, rijksmuseum.nl/en/collection/art-explorer. Accessed 9 April 2025.

– Rombach, Robin, et al. "High-Resolution Image Synthesis with Latent Diffusion Models." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, arxiv.org/pdf/2112.10752 Accessed 12 June 2025

– Rutherford, Ananda. "Working with Machine Learning: Research Reflection." *MuseumsxMachinesxMe*, edited by Susan pui san lok and Mark Miller, Tate Publishing, 2024.

– Schuhmann, Christoph, and Roman Beaumont. "LAION-Aesthetics." *LAION*, 2022, laion.ai/blog/laion-aesthetics/ Accessed 21 February 2025.

– Schuhmann, Christoph, et al. "Laion5b: An Open Large-Scale Dataset for Training Next Generation Image-Text Models." *arXiv*, 2022, arxiv.org/abs/2210.08402 Accessed 12 June 2025.

– Scott, Carol, et al. "User Value of Museums and Galleries: A Critical View of the Literature." Arts and Humanities Research Council, 8 July 2014.

– Sluis, Katrina. "Photography Must Be Curated! Part Four: Survival of the Fittest Image". Still Searching: An Online Platform for Photographic Discourse. Fotomuseum Winterthur, Switzerland. 2019. https://sites.rutgers.edu/critical-ai/wp-content/uploads/sites/586/2021/10/Sluis_2019_Survival-of-the-Fittest-Image.pdf Accessed 12 June 2025

– Sluis, Katrina, and Daniel Palmer. "The Automation of Style: Seeing Photographically in Generative AI." *Media Theory*, vol. 8, no. 1, 2023, p. 160.

– Song, Yehwan, and Joava Krysa. "Newly Formed City, Helsinki Biennale." *DV Studies*, 13 June 2023, dvstudies.net/2023/06/13/newly-formed-city-ai-curation-helsinki-biennial/. Accessed 17 Feb. 2025.

– Soydaner, Derya, and Johan Wagemans. "Unveiling the Factors of Aesthetic Preferences with Explainable AI." *British Journal of Psychology*, 2024, pp. 1-35, doi.org/10.1111/bjop.12707. Accessed 25 July 2025

– Steyerl, Hito. "In Defense of the Poor Image." *e-flux*, Nov. 2009. worker01.e-flux.com/pdf/article_94.pdf. Accessed 12 June 2025.

– Steyerl, Hito. *Medium Hot: Images in the Age of Heat*. Verso, 2025.

– Styles, Eleanor Brooke. "Tate Worlds Art and Artifacts Reimagined in Minecraft." *Advances in Archaeological Practice*, vol. 4, no. 3, 2016, p. 413.

– Thiel, Sonja, and Jonathan Bernhardt, editors. *AI in Museums: Reflections, Perspective and Interpretations*. Transcript Verlag, 2023, library.oapen.org/bitstream/id/3ecbc4ec-2dac-4e05-881a-10414c20f7f2/9783839467107.pdf. Accessed 12 June 2025

– Zouli, Ioanna. "Digital Tate: The Use of Video and the Construction of Audiences." PhD dissertation, London South Bank University, 2018. https://openresearch.lsbu.ac.uk/item/8683y Accessed 12 June 2025.

Biography

Sami P. Itävuori (he/they) is a London-based researcher, curator and cultural programmer with a specific interest in advanced technologies, audio-visual cultures and contemporary museum practices. Their practice is informed by community-centring approaches that promote skill-sharing, self-organization and alternative modes of making and art. They are on the board nomination committee of Anrikningsverket/Norbergfestival and are a PhD student at London South Bank University's Centre for the Study of the Networked Image, the Royal College of Art and Tate.