



Faculty of Design

2023

## How to Read a Generative AI Image System: Diffusion models as a techno-social entanglement

Salvaggio, Eryk

---

### Suggested citation:

Salvaggio, Eryk (2023) How to Read a Generative AI Image System: Diffusion models as a techno-social entanglement. In: Proceedings of Relating Systems Thinking and Design Volume: RSD12, 06-20 Oct 2023. Available at <https://openresearch.ocadu.ca/id/eprint/4921/>

*Open Research is a publicly accessible, curated repository for the preservation and dissemination of scholarly and creative output of the OCAD University community. Material in Open Research is open access and made available via the consent of the author and/or rights holder on a non-exclusive basis.*

*The OCAD University Library is committed to accessibility as outlined in the [Ontario Human Rights Code](#) and the [Accessibility for Ontarians with Disabilities Act \(AODA\)](#) and is working to improve accessibility of the Open Research Repository collection. If you require an accessible version of a repository item contact us at [repository@ocadu.ca](mailto:repository@ocadu.ca).*



**Relating Systems Thinking and Design  
(RSD12) Symposium | October 6–20, 2023**

## **How to Read a Generative AI Image System: Diffusion models as a techno-social entanglement**

**Eryk Salvaggio**

Generative media synthesis tools have quickly gathered the public's attention through products such as ChatGPT for text, or DALL-E 2, Midjourney and Stable Diffusion for images. However, the design of these systems is typically obscured through interfaces (buttons labelled "imagine" or "dream") and through the misleading label of "intelligence," with commentators likening these systems' behaviours to human creativity and ingenuity. Until now, the functions of these systems have focused on machine learning white papers, narrowly addressing internal technical processes. As policymakers, educators and the public grapple with these black boxes, this paper offers a systems-level analysis to clarify the entanglement of these technical systems within a broader context of data collection practices, generative models, user interfaces, generated and source images, and the broader media and cultural spheres in which they circulate. Ecological impacts and human labour concerns are also acknowledged. This paper maps out a systemic analysis of generative AI using a particular AI image generation system, Stable Diffusion, intended as a model and means to provide a common language for discussing and addressing these entanglements. Revealing the structures and relationships between the "systems within AI systems" is a means to engage with ethical controversies and techno-social possibilities more thoughtfully.

KEYWORDS: Artificial Intelligence, Generative AI, Diffusion Models, AI Ethics

RSD TOPIC(S): Cases & Practice, Mapping & Modelling, Society & Culture, Sociotechnical Systems

---

## Introduction

The rise of data-driven online services such as social media, text and photo sharing websites have enabled infrastructures of data capture and storage at unprecedented scales. This escalation of storage capacity in massive data centres has, in turn, contributed to the development of tools to manage and sort it. Information from millions of social projects – from building an online encyclopaedia (Gabrilovich, 2014) to gathering personal photos to share with family members (Karras, 2021) – have become datasets for training neural networks for pattern recognition and prediction tasks.

In 2022, OpenAI launched DALL-E 2 (Ramesh, 2022), described as “an AI system that can create realistic images and art from a description in natural language” (OpenAI 2022). This was followed by Stable Diffusion and Midjourney. These systems produce photorealistic, artistic quality images in a variety of styles in response to user prompts (texts describing images in natural language), with images that are closely aligned with the user’s description. Alongside this shift in image production was a parallel series of text-based AI tools, such as OpenAI’s GPT series of Large Language Models. The rise of these data-driven, algorithmic pattern-recognition systems quickly led to media and public speculation about sentience in AI systems (Tiku, 2022), protests from creative professionals such as writers and illustrators (Poniewozik, 2023) and calls for bans in classrooms and universities (Jimenez, 2023).

In this paper, the term “artificial intelligence” is best understood as an umbrella for a research field that aims to automate human decision-making based on pattern recognition within datasets. There is no “artificial intelligence” system per se, but a series of distinct systems brought together under the AI umbrella. It is more precise to speak about Large Language Models or Diffusion systems, each of which can contain distinct approaches, models, and methods. This paper focuses on a single output—synthetic images made by Diffusion models, such as Stable Diffusion and

Midjourney—and looks to the underlying system through the lens of relationships rather than as isolated components of a singular, dedicated system. As media, these systems are entangled with the ways humans manufacture and produce meaning: they are inscribed with, reproduce, and recirculate forms of social and cultural values. When these social values connect to aspects of the technical system, we can describe them as techno-social entanglements.

While examining these systems from a technical perspective is useful, it is an incomplete view of the system if the cultural and social systems of images and text are set aside. But we could compile an entire bestiary of entangled relationships with these systems. As a proposed taxonomy of AI entanglements, I examine the generated image as the result of at least four distinct systems, broadly defined as:

1. **Data:** Images and text in the dataset and their collection and categorising methods.
2. **Interface:** Prompts or other tools to access the data layer.
3. **Image:** The result of the data, user, and interface in interaction.
4. **Media:** Distributed social-cultural ideas of images and their meaning.

Each of these systems is embedded within even larger systems of human labour and environmental resources. Each subsystem has its own organising logic, methodologies, and concept of a user or outcome. They interact differently with the world beyond them, responding to and exerting unique pressures. This paper is organised to respond to each layer of the system through the lens of these entanglements, raising ethical concerns and proposing even further extensions of these systems.

As a study of socio-technical entanglements, this paper is both tangled up itself and yet permanently incomplete: it is intended only as a rough assemblage of lenses for organising the analysis of Artificial Intelligence image models, drawing from cybernetics and system design frameworks. Likewise, as a paper meant for designers and policymakers, it sets a boundary at the input of policy and design perspectives.

## **Data: sources, collection, and models**

Data is the core of these generative systems (Ramesh, 2021). Historically and practically, data is what allowed these systems to be conceived and to function. On a historical level, theories of knowledge built on data collection, analysis, hypothesis and prediction set the stage for this frame of intelligence. On a practical level, generative systems are automated pattern recognition systems—analysing datasets, finding patterns, categorising those patterns and then reproducing them. Drawing from data analysis best practices, we can understand the output of these systems best when we examine the biases, limits, and structures of the underlying dataset. An analysis of these systems begins with questions adapted from Gebru (2018):

1. What is the data?
2. What is included and excluded?
3. Where does that data come from?
4. How was this data selected and processed?

To better understand the data sub-system, we can begin by breaking data down into its sources, how it is collected and stored, and what the model does with it.

### **Data sources and their collection**

The data used by Stable Diffusion is not limited to images but billions of image-and-text pairs, that is, images and the text associated with them: captions, metadata, or alt-text for accessibility features. This data is chiefly extracted from the World Wide Web via the Common Crawl public web archive (Schuhmann, 2022), but some services may draw from other public or private data sets.

When drawn from the World Wide Web, human users have uploaded photographs, illustrations, and other visual data (which can include images as diverse as medical records, advertisements, diagrams, and more). These images, paired with human-generated descriptions (5), are recorded as a set. In many cases, archives and their machine-generated descriptions are also incorporated into the dataset.

One such collection of image-text pairs is LAION-5B, which stores five billion “text-and-image pairs,” the URLs of images alongside related text descriptions

(Schuhmann, 2022). LAION-5B is not an archive of the images themselves. Nonetheless, the model used by Stable Diffusion draws from these 5 billion data points to arrive at an image.

### **Diffusion models**

Diffusion models work by removing information from images until they arrive at a frame of pure noise or digital static. It begins by removing small amounts of information while analysing the distribution (or diffusion) of noise. It keeps a record in the form of a mathematical formula or algorithm for how that noise is repaired, which creates a pathway back to the original image (Sohl-Dickstein, 2015).

This information is then categorised according to the text associated with that image. Categorical clusters contain pattern-finding algorithms associated with specific frames of static. The model does not draw from the images themselves but from information about how the subjects of those images break apart. Gaussian noise follows a pattern where it tends to cluster around the densest part of an image. This reduces that image to core shapes. When the image is completely obliterated, it has learned to trace a frame of pure static back to that picture.

Later, when the model is activated by the user through the interface, the model generates a frame of random static. The user's text is used to steer the model toward repairing that noise according to paths associated with specific categories of images.

### **Entanglements of data and socio-cultural systems**

Expanding the boundaries of the data level of the system is useful in examining where data comes from, and extends the technical system into social, cultural, legal and political systems. Social and cultural systems permeate the LAION-5B dataset. It collects not only benign images of cats and butterflies but also a range of pornographic images, white supremacist and racist memes, and other stereotyping content (Birhane, 2021). This is a key result of human decisions to automate assembly and organisation. The LAION-5B white paper confirms this, noting that the authors are "against any applications in deployed systems without carefully investigating behaviour and possible biases of models trained on LAION-5B" (Schuhmann, 2022).

Nonetheless, data is itself entangled with a model of image-making which relies on central tendencies. Attempts to localise training data to specific regions of the world continue to create stereotypes but in more localised forms. In one example, a version of Stable Diffusion trained on local data from India merely created more specific stereotypes of Indian weddings (Trends Desk, 2023).

### **Entanglements of data and legal systems**

There are also questions about whether this data was ethically collected. Significant portions of the dataset are being contested as a violation of copyright (Appel, 2023). These lawsuits view the data in the model as a derivative of copyrighted images. Defendants see the data as the result of legitimate research and the model as removed from that data collection. Such research is argued to be acceptable under fair use of that data based solely on an input-output model of these systems. But, the broader ramification of that finding can be contested when viewed from a broader systems lens, which is explored in the labour section of this paper.

### **Entanglements of data and ecological systems**

AI image generation systems maps should consider systems of resource extraction. The first approach to resource extraction is the data commons of the world wide web, which is extracted through Common Crawl into datasets. However, there is also an environmental extraction: it is estimated that 700,000 litres of water were used to cool the data centres training GPT3 (Li, 2023) and that Large Language Models can emit the Carbon Dioxide of five cars over their entire lifetime during the training alone (Strubel, 2019). Training Stable Diffusion 1.4 used an estimated 11.25 metric tons of carbon dioxide, according to Hugging Face's model card (2023), while generating a single image is estimated to require the same amount of electric power as a full iPhone charge (Luccioni & Jernite, 2023).

There are also physical resources involved in assembling GPUs, and the digital infrastructures that link users to these models. Literature examining the reliance on natural resources from these nascent technologies remains sparse, but should be centred in systems maps of generative systems.

## **Interfaces: Steering users through myths**

The interface layer of the image synthesis system might include the following:

- A method of entering a request for an image—typically a prompt, though other modes, such as image-to-image style transfer, is possible.
- A way of displaying the resulting image.
- System-level monitoring to prevent certain images from being displayed.
- Design cues and affordances that steer the user into an understanding of the system.

A user's prompt is a caption for an image that doesn't exist yet. Diffusion models search the noise in a png or jpg, and identify possible paths back to an image in the category associated with the text. The model has gathered numerous pixel constellations from the process of dissolving training images into noise, with additional autoencoding for increased detail and resolution (Rombach, 2022).

These objects contain a generalised, ideal form of the images they decompose: when prompted for an apple, it can pull information about all the apples it has previously encountered and apply effective denoising processes based on previous examples. The user's prompt generates a new frame of randomised noise. The model begins to remove noise from this frame toward the image described by the prompt as if the prompt were a caption. If the noise is random, it's unlikely—but not impossible—that it could trace random noise back to any specific apple that it has ever seen. It will reverse the noise by finding associated pathways through the digital noise, referencing an abstract composite from all possible apple shapes.

The images it produces are not composites per se. They draw from a space that is a mathematical representation of many images blended under a common category (accessible through the tokenisation of text). I suggest that this possibility space should be understood as a composited space: an algorithmic representation of all possible configurations of images in that category. While these models are frequently called generative, they might also be considered reductive, as the model must remove digital noise from an image until it arrives at information (the image).

## **Entanglements of prompts and models**

This reliance on limitation and reduction to category means these images are, by nature, constrained to the range of possibilities based on image-text information compiled in the data layer of the system. This contains socially constructed stereotypes of people and situations, which are further constrained by the technique of noise reduction, always in a direction steered by datasets. The interface layer begins at this boundary of user prompt, steering the model, wherein the system's mechanics activate various mathematical weights and biases, which emphasise the most common examples in the dataset over the least common.

Crucially, the words themselves are translated into tokens: fragments of human language, reduced to bits, which, stringed together, serve as guideposts to clusters from which the system may reference in its image.

## **Entanglements of Technical and Social Biases**

Human and machine bias becomes entangled. Human captions may include invisible categories: for example, wedding is commonly used to describe a straight couple getting married, as opposed to straight wedding. But when a gay couple gets married, it is often described as a gay wedding (Cain, 2018). There is a tendency among captions uploaded online to share this kind of normative bias. This is taken into these categories and reinforced when they generate new images that follow the same rules: when wedding generates heterosexual pairings, because the images categorised under wedding are primarily heterosexual. Users who uncritically reproduce this bias in their prompts will create images that reinforce and recirculate these biases, unless systems intervene.

## **Entanglements of creativity and appropriation**

Just as the data layer and the interface layer reinforce the concentration of social and cultural stereotypes across two distinct processes, the interface layer can also extend the use of art and illustration data by allowing users to cite specific artists' names in their prompts. Directly typing an artist's name for an imitation of their style is a common use case for these tools, with many viral examples re-imagining certain movies in a particular film director's visual style, such as mixing quirky indie director Wes

Anderson with the horror film *The Shining*, or the mystical psychedelia of director Alejandro Jodorowsky with the Disney cyberpunk *Tron* franchise. An early commercial foray into this space, a Coca-Cola advertisement, was set in an art museum, where the art was “remixed” to include a bottle of Coca-Cola, seemingly celebrating this spirit of appropriation as creativity. While satirising popular culture is nothing new, artists have started to take to the press with stories of their own work becoming lost in the sea of generated images online. The artist Greg Rutkowski has claimed that his work was used in prompts so often that his own work does not appear in search results (Heikkilä, 2022). Likewise, the top search result on Google for the painter Edward Hopper is an AI-generated imitator as of May 2023 (Lee, 2023).

### **Entanglements of design and imagination**

Design cues help users make sense of these technologies by creating a story about how they work and how to use them. These interfaces have buttons labelled “dream” (Stable Diffusion’s Dream Studio) or “imagine” (Midjourney), which obscures the reality of what is happening behind these systems. This onboarding relies on the shorthand that the machine is an artist or that the machine is dreaming. These are metaphors and can hide aspects of the model and how it works, including those that might be objectionable to users.

Asking a machine to dream of an image romanticises the reassembly of image information from the corresponding category. Such imaginations become entangled with questions of sentience and can create robust frames that distort real understanding of these systems in policy and community spaces.

### **Entanglements with content moderation systems**

Finally, the interface is also a site of system-level interventions. The system intervenes in what comes back to the user at two stages: when the prompt leaves the interface to trigger the model and when the image is produced by the model. Midjourney blocks certain keywords from being entered into the model, aiming to limit the production of illegal or concerning content. OpenAI’s DALL-E 2 seems to insert random, diversifying words into prompts without the user’s knowledge in a bid to diversify the images that

return from the data, a way of tackling invisible categories (Offert, 2022). DALL-E 2 will return an image of a black professor, for example, while the data most likely would not. Such interventions also concern questions of power and transparency. Content Moderation Systems are entangled in cultural and social biases. As I have previously shown, the censorship of “women kissing” by OpenAI’s DALL-E 2, while “men kissing” was acceptable, is a cultural decision (Salvaggio, 2023). In DALL-E 3, the user prompt is intercepted and transformed via “shadow prompting” (Salvaggio, 2023c), where words trigger unrelated tokens in order to diversify image requests or steer the user away from offensive content. Hugging Face’s Dream Studio uses machine learning to detect whether an image contains violent or pornographic content and (unreliably) applies a blur filter to offending images.

### **Images: Data, users and interfaces**

The dataset is present in an image, visible in what is present and absent in any given selection. Previous writing explores the image as an infographic or data visualisation in greater detail (Salvaggio, 2023) and will only be summarised here.

In raw form, generated images are visual representations of the category assignments in the underlying dataset. The image is its own entanglement of words, data, and internet culture, drawing specifically from the English language web’s consensus labelling of the images it relies on. The image is a visual system representing these relationships and is inscribed primarily by the arrangement of these categories and the contents paired with language.

### **Media: Interactions and circulations of generated images**

AI images recirculate through the media sphere. If they are drawn from stereotypes—whether it is a stereotype of people or kitchens or apples—they will recirculate those stereotypes. Data that goes into an AI image, if recollected to train future models, creates a self-sustaining feedback loop.

Users bring specific cultural values and understandings to these images. Images tend to show a Western bias, as seen in the naturalisation of smiles in a series of generated “selfies” (Gurfinkle, 2023). Likewise, cultural representations of spaces such as “kitchens”

would most likely reveal Western kitchens. For this reason, media systems are part of the expanded boundary of image synthesis systems. The media system is part of the production of the data in the data layer, shapes the design of the interface layer, informs the way users write and categorise prompts on the user end, and read and confirm the resulting images on the output end.

### **Entanglements of media and power**

Media is where images and ideas are distributed and circulated, encoded and decoded. Meaning is produced and interpreted through these structures. While the World Wide Web provided a temporary participatory environment in which the public could respond to the state and corporate actors, generative interfaces are shaped by corporations, who have tremendous power over media (through advertising and public relations agencies) and over the state (through lobbyists and political donations). While the participatory internet allows for imperfect deliberation by the public, images generated by AI systems, mostly through closed models, merely reproduce an algorithmically determined consensus. As of 2023, there is no public response to a closed loop of generative media; the dataset becomes a de facto status quo.

Increasingly, that is itself mediated by an algorithmic layer, which dictates which content people see (and which they do not). The interaction of image algorithms and media algorithms, all taking place within a social sphere, is another important entanglement to examine critically (O'Neill, 2017).

### **Entanglements of consumption and production**

A key question about AI-generated images is who these images are for. The audiences for these images vary: they may be understood as images, exchanged on social media or sold as Non-Fungible Tokens in online marketplaces. They may also be considered interactive media, where the process of prompting these images is itself a form of entertainment—a blend of video game and self-directed storytelling. The architecture of these services suggests much of the latter: the images are resistant to copyright because they are made by machines (US Copyright Office, 2023).

Midjourney cultivates a community where every generation is shared to its Discord servers (a public chat feed) unless users pay steep monthly fees. This suggests that the generation of images is not merely a tool in the artistic process, such as Photoshop, but part of a distinct media form in which creation and consumption are entangled, akin to social media (Filippas, 2021) and video games (Burgess, 2022). This form can be solitary – creating images of your own desire and liking – or connected to communities of viewers. In either case, it presents shifts in the relationship between consumer and producer from that of traditional photography and illustration.

### **Entanglements of images and copyright**

Cultural concepts around generated images shape the way they are interpreted and understood. There is some philosophical debate about whether these systems produce artworks or mere statistical correlations and how to define the difference (Hertzmann, 2018). The idea that machines can generate art is simultaneously endorsed and rejected by the US Copyright Office (2023), which holds that the machine has rendered these images from statistics and that the human intervention in their production is not substantial enough to merit protection unless combined with human-authored creative input (2). This inadvertently attributes artistry to the machine, whereas the *intent* of the process is understood to be set by human users. One pushes the “dream” button to generate images focused on certain aspects of the dataset through prompting with Stable Diffusion. Another pushes the shutter of a camera whilst pointing it in a focused direction through the lens. Yet, the US copyright office has never held that the camera is the photographer (3). Important entanglements emerge between concept and expression in the legal domain: while ideas are not protected, expression is. How this is navigated by the legal system will require an ongoing series of arguments between creative users of AI systems and the copyright office.

## **Entanglements of human labour**

AI image systems are falsely described as autonomous. In fact, humans are engaged in nearly every layer of these systems. Human labour enters at the data level, including data cleaners who prepare and label large image datasets or identify content moderation issues (Roberts, 2021). While OpenAI contrasts its model from ImageNet by using Image-Text pairs that were “already publicly available on the Internet,” (OpenAI, n.d. B), this is evidence of the presence, not absence, of human labour: it relied on images uploaded, labelled, and shared online by human beings. Boundaries of an AI system also include illustrators and artists whose work is analysed in the data layer of these systems; without their collective effort and contributions – voluntary or not – these datasets could not be trained, and these tools could not exist. New industries are emerging to accommodate and protect artists, from online tools to opt out of datasets (see spawning.ai) or invisibly alter images in ways that make learning from them impossible (Shan, 2023).

A systems view of AI image generation should account for who was paid for that labour, who was not, and how much. It would track the economics of AI companies and media companies and the influence that they have on shaping our understanding of how these systems work and what they do.

Human labour is also entangled in the interface layer. This includes hiring, supervision, and business decisions affecting management, designers and programmers building interfaces and algorithms, and the incentives that guide them. Likewise, the images themselves and how they circulate enter a precarious market of illustrations and for-hire photography, creating interventions and disruptions in the existing media industry, where image and media layers are interconnected.

Designers make decisions about the operation and structure of these systems. This includes the decision to deploy technologies regardless of explainable outcomes. The unexplainable outcomes of AI systems should not shield designers and companies from accountability for deploying unaccountable systems. This framework proposes that decisions about the AI system are intentional and offers a framework for understanding such accountability.

## Diagramming an AI system

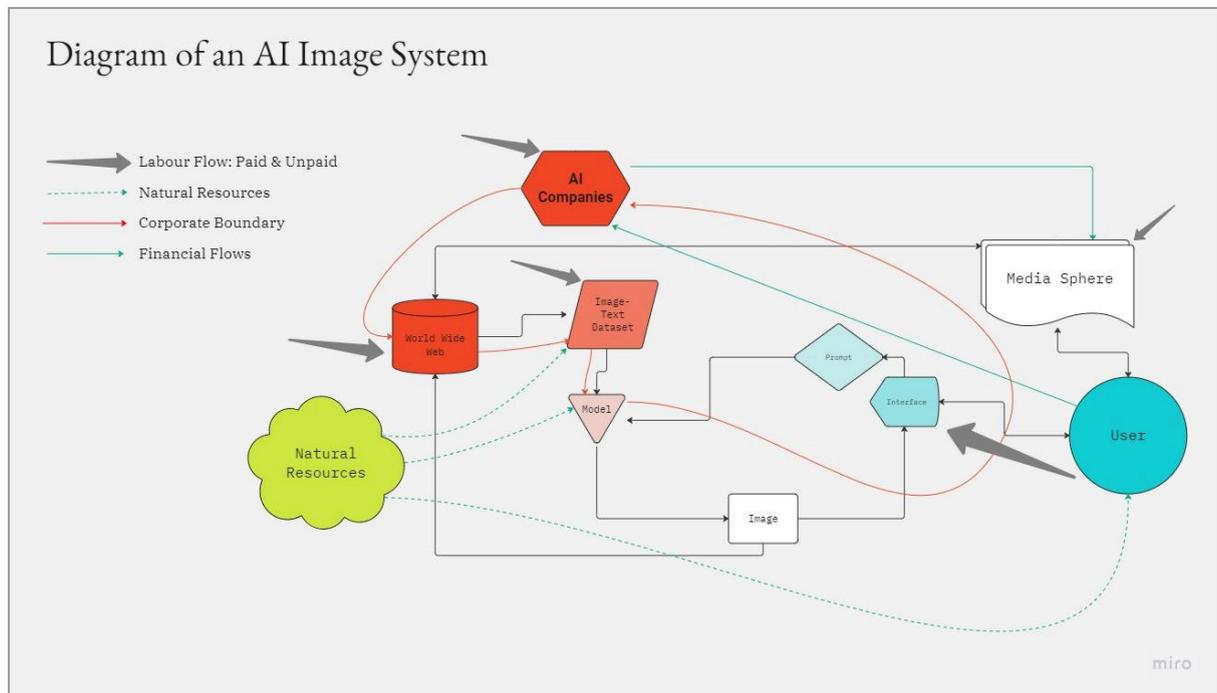


Figure 1: A systemic diagram of an AI Image System depicting connections between technological, social, natural, corporate, media and natural systems and the nature of information exchange between them.

The systemic map in Figure 1 is intended as a tool for generating conversation and questions among policymakers, designers, communities or classrooms. It represents a framework on which concerns and opportunities in generative AI might be identified. Red is meant to highlight areas of corporate enclosure: for example, operators such as Stable Diffusion rely on the World Wide Web to mine data in ways that very much assume it is theirs to use. They also rely on the image-text dataset (such as LAION-5B) to train the model. Within this red enclosure is the interaction between the prompt and interface, which is the user's point of entry into the system. Meanwhile, natural resources are flowing into the model, the dataset collection, editorial decisions such as content guidelines and risk mitigation, and to the user. Beyond this corporate enclosure is the image itself, which is deemed public domain under current copyright law.

Meanwhile, the image is engaged in a loop with the interface, as the user can generate variations on the resulting image or use a prompt to create new images. This is a point of user control and interaction. The image can also be recirculated into the World Wide Web, where it interacts with the media sphere (such as search engines or social media networks). The image also risks being included in future training data. In the meantime, the media sphere is shaped and shapes users, with corporate contributions to the media sphere (in the form of advertising and public relations). There is an exercise in power between AI companies and the media sphere, which may influence system users, but there is likewise an opportunity for users to steer the media sphere to shape their own uses (as we have seen in the case of organised labour) (Salvaggio, 2023b).

Focal points for labour are also presented: whose labour is involved, and how it is compensated? Through the discussion or production of similar maps, affected or interested communities may talk through the many layers of the system, creating opportunities to better respond or mitigate their potential for harm.

For example, significant conversation around the production of generated images centres on copyright concerns about the images they produce. Through this model, we can also clearly see that there are concerns about what the models train on and how that data is obtained. The image is seen not in isolation but as part of a system of media distribution, which suggests other harms (such as artists' names being associated with images they did not make). In sum, the systemic map of AI images creates multiple points of entry for a system that, when focused entirely on images and users, is commonly understood as closed.

## Discussion

Stafford Beer wrote that “the purpose of a system is what it does” (2002), and it is in that spirit that I aimed to catalogue the entanglements that recursively shape what these systems do. No single conference paper can examine the full and expansive boundaries (or boundarylessness) of a technical system’s entanglement with systems of human culture. Instead, the paper suggests that systems mapping and systems thinking are essential in determining which lenses may be applied to interactions when making decisions about the design, deployment, and regulation of generative AI systems. Each entanglement is a potential starting point for the analysis of power in these systems, which are too often abstracted and then attributed to an unknowable “black box” algorithm.

By focusing on socio-technical entanglements, we inevitably expand a system well beyond its technical boundaries or relying on passive examination of social impact. Instead, we examine contact points where systems exchange information and adapt to one another to different degrees. The systemic view of an AI system reveals mechanisms for tracing harms, control, and accountability and describes how power is distributed or concentrated throughout these entanglements. Reading a system through its entanglements allows new questions to emerge, including key questions about where policymakers, designers, and communities may intervene to shape these systems.

For policymakers, it is the balance of power between these systems that suggests avenues and approaches to crafting regulation or other protections. For designers of AI systems, it is an essential lens for configuring ethical and responsible design and deployment of these systems. Optimistically, a better index of the entanglements of AI systems can help to structure better, more equitable systems, especially when paired with a critical view of who gets to make the decisions within these systems and when they should be deployed.

AI models are discussed too narrowly within specific disciplines to provide a complete understanding of their impacts - harms, but also potentials. While drawing boundaries for systems is a subjective exercise, cybernetics and systems design lenses offer unique, reflexive tools for redefining those boundaries within generative AI. The reach of

artificial intelligence and its definition as a multi-modal, multidisciplinary cyber-physical systems, human labour and environmental resources. Shifting the scale of focus into each of these entanglements, rather than viewing specific slices of the system in isolation, is helpful in examining AI. It is a tool for identifying the broader implications of these images, identifying accountability among actors, establishing broader responsibilities to communities and users, and identifying opportunities for inclusion in the decision-making system, which demands a coordinated multi-modal, multidisciplinary response in research, design, and analysis. It suggests a return to the founding science behind Artificial Intelligence, rooted in an understanding of complex social drivers that is lost in today's AI research. It suggests an opportunity for contemporary cybernetics, committed to expanding its touchstones beyond Western viewpoints, to more broadly examine and critique systems which increasingly define, shape, and enact power. The AI Decolonial Manifesto, written by "two dozen experts in AI, social sciences, humanities, human rights, from different geographic regions, racial gender, age, class, and other intersectionalities," notes that "the social and the technical are interwoven, and technologies have immaterial as well as material impacts over specific gendered, racialised bodies and territories. Decoloniality rejects the divorcing of the material and immaterial, of feeling from being, knowing, doing or living" (Manifesto.ai, 2022).

In this spirit, the systems map aims to complicate our isolated understanding of "an AI system" to see it as necessarily embedded within a set of social, cultural, and material entanglements. This approach is a starting point for understanding power and is thus incomplete. There is no such thing as an "AI-generated image" from diffusion models, only images produced through intertwined systems of data, interface, images, and media.

## References

1. Appel, G., Neelbauer, J., & Schweidel, D. A. (2023, April 7). Generative AI has an intellectual property problem. *Harvard Business Review*.  
<https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>
2. Beer, S. (2002). What is cybernetics? *Kybernetes. The International Journal of Cybernetics, Systems and Management Sciences*, 31(2), 209–219.  
doi:10.1108/03684920210417283
3. Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). *Multimodal datasets: misogyny, pornography, and malignant stereotypes*. In arXiv [cs.CY].  
<http://arxiv.org/abs/2110.01963>
4. Burgess, J., & Jones, C. (2022). Exploring emergent co-creative narrative in a strategy video game brand. *Journal of Media Business Studies*, 19(3), 185–202.  
doi:10.1080/16522354.2021.1959705
5. Cain, S. (2018, August 16). A Gay Wedding Is a Wedding. Just a Wedding. *The New York Times*.  
<https://www.nytimes.com/2018/08/16/fashion/weddings/a-gay-wedding-is-a-wedding-just-a-wedding.html>
6. Filippas, A., & Horton, J. (2021). *The production and consumption of social media*. Cambridge, MA: National Bureau of Economic Research.
7. Gabrilovich, E., & Markovitch, S. (2014). Wikipedia-based semantic interpretation for natural language processing. *The Journal of Artificial Intelligence Research*, 34, 443-498. <https://doi.org/10.1613/jair.2669>
8. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2018). *Datasheets for Datasets*.  
<https://doi.org/10.48550/ARXIV.1803.09010>
9. Gurfinkle, J. (2023). AI and the American Smile. *Medium*. 26 March.  
[HTTPS://MEDIUM.COM/@SOCIALCREATURE/AI-AND-THE-AMERICAN-SMILE-76D23A0FBFAF](https://medium.com/@SOCIALCREATURE/AI-AND-THE-AMERICAN-SMILE-76D23A0FBFAF).
10. Heck, Marina Carmago. "The ideological dimension of media messages." In *Culture, media, language: working papers in cultural studies*, 1972-1979. London: Routledge, 1992.

11. Heikkilä, M. (2022, September 16). This artist is dominating AI-generated art. And he's not happy about it. *MIT Technology Review*.  
<https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>
12. Hertzmann, A. (2018). Can Computers Create Art? *Arts*, 7(2), 18.  
doi:10.3390/arts7020018
13. Hugging Face (n.d.). (2023). Model Card: CompVis/stable-diffusion-v1-4. Retrieved May 29, 2023, from Huggingface.co website:  
<https://huggingface.co/CompVis/stable-diffusion-v1-4>
14. Jimenez, K. (2023, January 30). "This shouldn't be a surprise" The education community shares mixed reactions to ChatGPT. USA Today.  
<https://www.usatoday.com/story/news/education/2023/01/30/chatgpt-going-banned-teachers-sound-alarm-new-ai-tech/11069593002/>
15. Karras, T., Laine, S., & Aila, T. (2021). A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4217-4228. <https://doi.org/10.1109/TPAMI.2020.2970919>
16. Lee, PN [@PennyNevilleLee]. (2023, May 3). If you don't think visual AI is a problem, this is what comes up when you @Google Edward Hopper.  
pic.twitter.com/w1x4yku76g. Twitter.  
<https://twitter.com/PennyNevilleLee/status/1653791716682477569>
17. Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). *Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models*. doi:10.48550/ARXIV.2304.03271
18. Luccioni, A.S., & Jernite, Y. (2023). *Power Hungry Processing: Watts Driving the Cost of AI Deployment?* [doi.org/10.48550/arXiv.2311.16863](https://doi.org/10.48550/arXiv.2311.16863)
19. Manyfesto.ai (Aarathi Krishnan, Angie Abdilla, A Jung Moon, Carlos Affonso Souza, Chelle Adamson, Eileen M. Lach, Farah Ghazal, Jessica Fjeld, Jennyfer Taylor, John C. Havens, Malavika Jayaram, Monique Morrow, Nagla Rizk, Paola Ricaurte Quijano, R. Buse Çetin, Raja Chatila, Ravit Dotan, Sabelo Mhlambi, Sara Jordan & Sarita Rosenstock) (2022) *AI Decolonial Manyfesto*. manyfesto.ai.
20. Offert, F., & Phan, T. (2022). A sign that spells: DALL-E 2, invisual images and the racial politics of feature space. Retrieved from <http://arxiv.org/abs/2211.06323>
21. O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

22. OpenAI. (n.d.). Dall·e 2. *Openai.com*. Retrieved May 24, 2023, from <https://openai.com/product/dall-e-2>
23. OpenAI . (n.d.). (b) CLIP: Connecting text and images. (n.d.). Retrieved May 29, 2023, from Openai.com website: <https://openai.com/research/clip>
24. Poniewozik, J. (2023, May 10). TV's war with the robots is already here. *The New York Times*.  
<https://www.nytimes.com/2023/05/10/arts/television/writers-strike-artificial-intelligence.html>
25. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). *Hierarchical text-conditional image generation with CLIP latents*. Ithaca: Cornell University Library, arXiv.org.
26. Roberts, S. T. (2021). Your AI Is a Human. In *Your Computer Is on Fire* (pp. 51–70). The MIT Press.
27. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-resolution image synthesis with latent diffusion models*. (). Ithaca: Cornell University Library, arXiv.org.
28. Salvaggio, E. (2023). How to Read an AI Image: Toward a Media Studies Methodology for the Analysis of Synthetic Images. *IMAGE 37*, (pp. 82-99).  
<https://image-journal.de/how-to-read-an-ai-image/>
29. Salvaggio, E. (2023b). What the WGA Contract Tells Us about Workers Navigating AI. Tech Policy Press. September 28, 2023. <https://techpolicy.press/what-the-wga-contract-tells-us-about-workers-navigating-ai/>
30. Salvaggio, E. (2023c). Shining a Light on Shadow Prompting.  
<https://www.techpolicy.press/shining-a-light-on-shadow-prompting/>
31. Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk, R., & Jitsev, J. (2022). LAION-5B: An open large-scale dataset for training next generation image-text models. In arXiv [cs.CV].  
<http://arxiv.org/abs/2210.08402>
32. Shan, S., Cryan, J., Wenger, E., Zheng, H., Hanocka, R., & Zhao, B. Y. (2023). GLAZE: Protecting artists from style mimicry by text-to-image models.  
doi:10.48550/ARXIV.2302.04222

33. Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. Cornell University Library, arXiv.org.
34. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. doi:10.48550/ARXIV.1906.02243
35. Tiku, N. (2022). The Google engineer who thinks the company's AI has come to life. *The Washington Post*.  
<https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-le-moine/>
36. Trends Desk. 2023. AI-Generated Art of Indian Wedding Couples Sparks Debate. *The Indian Express*, January 3, 2023.  
<https://indianexpress.com/article/trending/trending-in-india/ai-generated-art-of-indian-wedding-couples-sparks-debate-8359294/>
37. US Copyright Office. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence. 2023,  
<https://www.copyright.gov/newsnet/2023/1004.html>

## **Author**

Eryk Salvaggio, Independent Researcher, [cyberneticforests.com](http://cyberneticforests.com),  
eryk.salvaggio@gmail.com