

# Creating AI Art Responsibly: A Field Guide for Artists

**How to cite this article:** Leibowicz, C. R.; Saltz, E., & Coleman L. (2021). Creating AI Art Responsibly: A Field Guide for Artists. *Diseña*, (19), Article.5. <https://doi.org/10.7764/disena.19.Article.5>

DISEÑA 19

AUGUST 2021

ISSN 0718-8447 (print)

2452-4298 (electronic)

COPYRIGHT: CC BY-SA 4.0 CL

## Project

Reception

NOV 02 2020

Acceptance

JUN 23 2021

 Traducción al español aquí

**Claire R. Leibowicz**

Partnership on AI

**Emily Saltz**

The New York Times

**Lia Coleman**

Rhode Island School of Design

Machine learning tools for generating synthetic media enable creative expression, but they can also result in content that misleads and causes harm. The *Responsible AI Art Field Guide* offers a starting point for designers, artists, and other makers on how to responsibly use AI techniques and in a careful manner. We suggest that artists and designers using AI situate their work within the broader context of responsible AI, attending to the potentially unintended harmful consequences of their work as understood in domains like information security, misinformation, the environment, copyright, and biased and appropriative synthetic media. First, we describe the broader dynamics of generative media to emphasize how artists and designers using AI exist within a field with complex societal characteristics. We then describe our project, a guide focused on four key checkpoints in the lifecycle of AI creation: (1) dataset, (2) model code, (3) training resources, and (4) publishing and attribution. Ultimately, we emphasize the importance for artists and designers using AI to consider these checkpoints and provocations as a starting point for building out a creative AI field, attentive to the societal impacts of their work.

#### Keywords

Synthetic media

AI art

Responsible AI

AI ethics

Generative media

**Claire R. Leibowicz**—BA in Psychology and Computer Science, Harvard University. Master in the Social Science of the Internet, University of Oxford (as a Clarendon Scholar). She is the Head of the AI and Media Integrity program at the Partnership on AI, a global multistakeholder nonprofit devoted to responsible AI. Under her leadership, the AI and Media Integrity team investigates the impact of emerging AI technology on digital media and online information. She is also a 2021 Journalism Fellow at *Tablet Magazine*, where she is exploring questions at the intersection of technology, society, and digital culture, and an incoming doctoral candidate at the Oxford Internet Institute. Her latest publications include, 'Encounters with Visual Misinformation and Labels Across Platforms: An Interview and Diary Study to Inform Ecosystem Approaches to Misinformation Interventions' (with E. Saltz and C. Wardle; *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Issue 340) and 'The Deepfake Detection Dilemma: A Multistakeholder Exploration of Adversarial Dynamics in Synthetic Media' (with A. Ovadya and S. McGregor; *Proceedings of the 2021 ACM Conference on Artificial Intelligence, Ethics, and Society*).

**Emily Saltz**—Master in Human-computer Interaction, Carnegie Mellon University. She is a UX Researcher studying media and misinformation, working with organizations like the Partnership on AI and First Draft. She led UX for The News Provenance Project at The New York Times, where she works as a UX researcher. Some of her work includes a collaboration on an AI-generated op-ed for author Oobah Butler on being catfished by AI (*The Independent*, 2021); explorations of text prediction software such as 'Human-Human Autocompletion' (presented at WordHack at Babycastles, 2020) and 'Super Sad Googles' (presented at Eyeo 2019); and 'Filter Bubble Roulette', a mobile VR experience to inhabit user-specific social media feeds (presented at The Tech Interactive in San Jose, 2018).

**Lia Coleman**—BSc in Computer Science, Massachusetts Institute of Technology. She is an artist, AI researcher, and educator. Adjunct Professor at Rhode Island School of Design, she teaches machine learning artwork. She is the author of 'Machines Have Eyes' (with A. Raina, M. Binnette, Y. Hu, D. Huang, Z. Davey, and Q. Li; in *Big Data, Big Design: Why Designers Should Care About Machine Learning*; Princeton Architectural Press, 2021), 'Art'ificial (with E. Lee; *Neocha Magazine*, 2020), and 'Flesh & Machine' (with E. Lee; *Neocha Magazine*, 2020). Some of her recent workshops and talks include 'How to Play Nice with Artificial Intelligence: Artist and AI Co-creation' (presented at Burg Giebichenstein University of Art and Design, 2021); 'A Field Guide to Making AI Art Responsibly' (presented at Art Machines: International Symposium on ML and Art), and 'How to Use AI for Your Art Responsibly' (presented at Mozilla Festival, 2020 and Gray Area, 2020).

## Creating AI Art Responsibly: A Field Guide for Artists

### Claire R. Leibowicz

Partnership on AI  
New York, USA  
[claire@partnershiponai.org](mailto:claire@partnershiponai.org)

### Emily Saltz

The New York Times  
New York, USA  
[essaltz@gmail.com](mailto:essaltz@gmail.com)

### Lia Coleman

Rhode Island School of Design  
Providence, USA  
[liailiad@gmail.com](mailto:liailiad@gmail.com)

---

## CREATING AI ART RESPONSIBLY: A FIELD GUIDE FOR ARTISTS

---

### Background of the problem

Artificial intelligence (AI) tools for generating media have become increasingly accessible (Lomas, 2020; Nicolaou, 2020), presenting the potential for synthetically generated media that misleads and causes harm. While deepfakes, or AI-generated images and videos, have captured the public's attention, even lower-tech techniques such as cheap fakes —like those frequently used in videos of politicians— can be used to alter perceptions of public figures and events (Chesney & Citron, 2018; Paris & Donovan, 2019).

However, the same AI tools provide artists and designers with a new creative field with unique affordances. Naoko Hara, for example, uses images drawn from her own animation work as data to generate art (Hara, 2020). Derrick Schultz's AI-generated art leverages images from floral illustrations (Figure 1). Synthetic media techniques have also been leveraged in film for privacy protection purposes (Li & Lyu, 2019). In the 2020 film 'Welcome to Chechnya', filmmakers protected the identities of subjects discussing LGBTQ+ experiences, allowing subjects to tell their stories safely (Rothkopf, 2020). Recently, AI artists have even used the technology to prove technology's effects on society, including those afforded by the synthetic media medium itself. In 2019, Bill Posters and Daniel Howe's installation 'Spectre' featured a deepfake video of Mark Zuckerberg to illuminate Facebook's influence on user behavior.



**Figure 4:** An output generated with CycleGAN using images of a cat and flower illustrations in the 'faces2flowers' project by Derrick Schultz (2019). The post features a link for readers to experiment with the source model themselves in the RunwayML platform. Source: Schultz, 2019. Image retrieved from <https://artificial-images.com/project/faces-to-flowers-machine-learning-portraits/>. Courtesy of Derrick Schultz.

As artists, designers, and other creators leverage AI technologies, specifically machine learning (ML) methods, it is crucial for them to understand the context and potential harms of their work through the lens of the broader synthetic media and AI research communities, and in turn, for those research communities to attend to the unique needs and insights of independent creators. While responsible AI guidelines for researchers and technology companies abound (Gebu et al., 2018; Mitchell et al., 2019; Raji & Yang, 2019), these typically focus on engineers and broader technical staff at large social media platforms and information and communication technology companies, rather than the unique goals and needs of

artists balancing creative expression and positive societal impact with the potential harms of their creations.

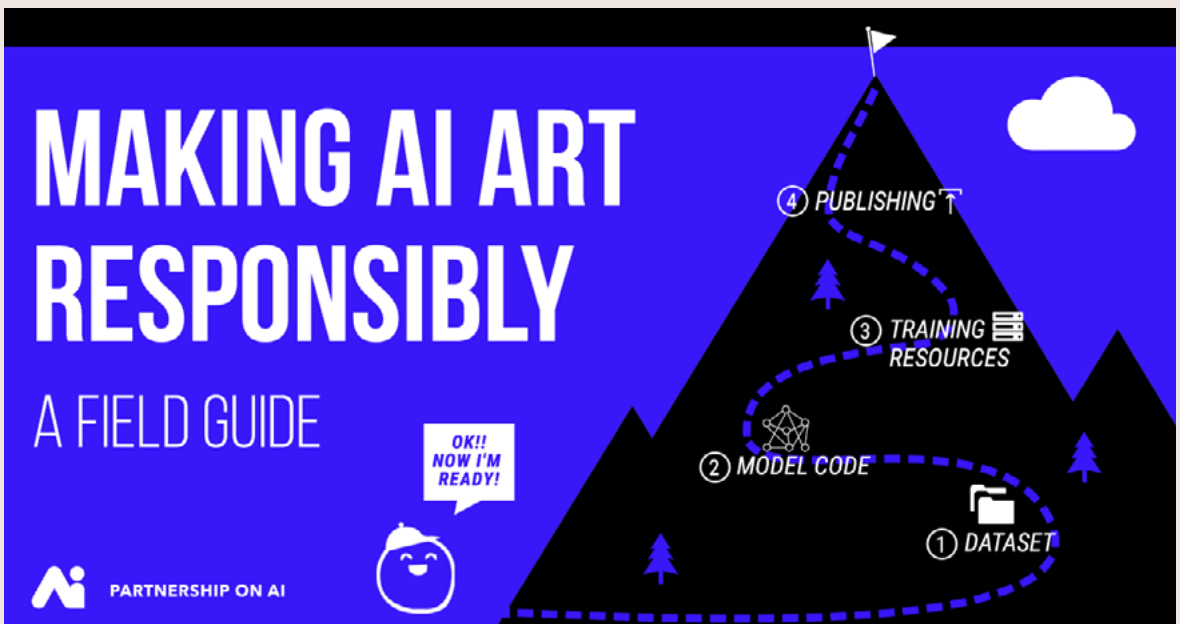
### Objectives

We, therefore, created a digital zine, the *Responsible AI Art Field Guide* (Figure 2), in order to equip AI artists and designers with insights on how to create AI art responsibly. It came out from the Partnership on AI's AI and Media Integrity Program that investigates the impact of emerging AI technology on digital media and online information (Saltz et al., 2020). The guide is structured around questions for artists and designers using AI to consider throughout the lifecycle of their creative practice to better situate their work within a responsible AI practice, blending insights from responsible AI practices with design tactics specifically intended for AI creators. It also offers emerging best practices drawn from the multidisciplinary experiences of AI artists, practitioners, researchers, and the scholarly community from fields such as communication, computer science, media forensics, sociology, and media studies.

**Figure 2:** Cover Image from the *Responsible AI Art Field Guide*, by Emily Saltz, Lia Coleman, and Claire R. Leibowicz (Partnership on AI, 2020). Screenshot: Claire R. Leibowicz, 2020. Retrieved from Medium: <https://medium.com/partnership-on-ai/a-field-guide-to-making-ai-art-responsibly-f7f4a5066ee>

### Conceptual framework

Our definition of 'responsible AI' considers both critiques of how corporations currently create and release AI products, as well as broader critiques of the corporate structures powering AI systems, including labor rights and worker equity (Diehm & Sindors, 2020; Rakova et al., 2020). In the field of AI-generated media,



ensuring responsible AI practice requires paying special attention to malicious use cases for the technology. Based on our work and collaboration with many organizations working on and around responsible AI and synthetic media at the Partnership on AI, we understand the harmful consequences of synthetic media to largely implicate the misinformation, information security, biased/appropriative content, copyright, and environmental domains. Of course, there are also non-malicious use cases for AI-generated media in art, design, and entertainment. We define AI art as all new works made with creative intent using techniques where computer programs access data and learn automatically from that data with minimal human intervention. The *Responsible AI Art Field Guide* seeks to compile emerging insights and best practices from both AI artists and the responsible AI and synthetic media research communities to help guide reflective processes for generating creative works and designing products with AI (while mitigating unintended consequences).

## Background

Most of the previous attention to the negative impacts of synthetic media has come from technology platforms focused on mitigating misinformation. Technology platforms like Facebook have invested heavily into preparing for the use of manipulated media to affect public opinion and spread misinformation — specifically focusing on policies and machine learning innovations to combat visual misinformation (Bickert, 2020; Dolhansky et al., 2020; Leibowicz, 2020; Roth & Achuthan, 2020). While most of the visual misinformation today is not AI-generated, AI is used most often for other malicious use cases that affect information security: to create images for non-consensual sexual exploitation. According to a 2019 report from Sensity, a visual threat intelligence firm, 96 percent of deepfakes online are pornographic in nature (Patrini, 2019).

In recent years, researchers have attempted to synthesize best practices for using ML datasets, such as ‘Datasheets for datasets’ (Gebru et al., 2018), ‘Model Cards for Model Reporting’ (Mitchell et al., 2019) and ‘ABOUT ML’ (Raji & Yang, 2019), in order to help people better address the bias inherent in machine learning models and identify how datasets may be skewed toward certain attributes. Additionally, organizations like the Algorithmic Justice League have emerged to communicate the potential harms and biases of AI technologies (Buolamwini, 2016). Much of the Algorithmic Justice League’s work has focused on the deployment of high-impact systems like facial recognition, and their four core principles include affirmative consent, meaningful transparency, continuous oversight and accountability, and actionable critique (Buolamwini, 2016). While all these elements can be applied to dual-risk AI systems to make them more responsible, including AI art, it will be vitally important to contextualize such responsible



AI principles with the goals and motivations of creators, designers, and artists using AI techniques. Unlike the negative environmental impact of synthesizing media using AI techniques, which is agnostic to the motivations of the creator, what responsible use of AI means to better mitigate bias and appropriate content, requires the sensitivity of the creator's motivations.

While many responsible AI resources offer technical details and rationales relevant to independent creators, they primarily focus on large-scale ML practices in industry and academia. As a result, these insights are unlikely to reach artists and designers using AI in creative contexts. While there are critical design frameworks relevant to AI design, such as the Design Justice Principles (Costanza-Chock, 2018), it may be unclear to a creator how a principle such as "we prioritize design's impact on the community over the intentions of the designer" should be applied in practice, specifically to the steps involved for creative AI, such as dataset creation, model code, training resources, and publishing.

There are some signs of a burgeoning literature exploring responsible AI considerations of creative practices; however, it should be expanded upon and incorporate practical recommendations for navigating the creative process and trade-offs inherent to such activities. This is what we sought to carry out with our *Responsible AI Art Field Guide*. The complexities of this task are apparent in works such as that of Lyons (2020), which offers a critical assessment of Kate Crawford and Trevor Paglen's 'Training Humans' exhibition. Their exhibition was intended to critique corporate practices for training computer vision systems and was presented alongside the valuable written work, 'Excavating AI: The Politics of Images in Machine Learning Training Sets' (Crawford & Paglen, 2019). Yet Lyons, a coauthor on the JAFFE dataset that the authors critique, points out that, in critiquing corporate practices such as using facial images and videos without consent, Paglen and Crawford themselves also reproduced and exhibited these same images without consent. Lyons refers to this oversight for their artistic use case as an 'ethical double-standard'. This debate highlights the need for further critical evaluation of the use of human data in AI systems for artistic and design purposes, as well as consideration of copyright and information security, including the use of personal images without informed consent and the stated terms of use for the datasets used.

---

## **DEVELOPMENT AND PARTIAL RESULTS**

### **Methodological framework**

The *Responsible AI Art Field Guide* was developed to address this gap and came out of multistakeholder input from AI artists, visual misinformation researchers, machine learning engineers, and policymakers. The input was initially sourced from a July 2020 talk with Gray Area, a cultural hub for art and technology in

San Francisco, and conversations with members in the Partnership on AI, a multistakeholder, global nonprofit organization devoted to responsible AI with over 100 Partner organizations from civil society, industry, media, and academia. The Gray Area talk included over 50 participants from a variety of backgrounds, including AI artists, AI researchers, technology policy stakeholders, and others. We enabled these participants to offer feedback on the first stage of the guide in a working document, which allowed us to hone the recommendations and checkpoints. Lia Coleman, an AI artist and designer, designed the guide as an online zine that walks a hypothetical AI artist through checkpoints for responsibly making AI art, centered on four elements: (1) dataset, (2) model code, (3) training resources, and (4) publishing and attribution. Thus, our methodological approach incorporates consultation with practitioners and scholars in the Partnership on AI and Gray Area communities, a review of design and art projects dealing with AI, a literature review pointing to a lack of responsible AI art guidelines benefiting the unique needs of artists, and the experience of two different AI artists, Lia Coleman and Emily Saltz, drawing upon their personal experiences to hone the guide. While much writing on responsible AI is centered on academic discourse or practitioners largely housed within technology and industry, we focused on the AI art and design community as a distinct audience for responsible AI that is underserved in the current literature (Rakova et al., 2020). Alongside question prompts, we include case examples to underscore the implications of AI creation on different societal features—including questions of ownership, environmental impact, attribution, explainability, and privacy.

### **Methodological strategy of the project**

The guide strategically avoids being prescriptive. When it comes to the nascent and evolving field of creative AI, many topics are subject to debate and should be built upon through practical trial and error in conversation with the rapidly changing visual misinformation field in industry and academia. Others in the field of synthetic media have taken this approach; for example, Twitter has recently described the need for its policy responses to visual misinformation to be ‘living documents’, emphasizing that they are “willing to update and adjust when [they] encounter new scenarios” (Twitter Safety, 2020). The same can be said of our AI art and design guide for navigating responsible creation. While we should aspire to airtight frameworks, at this stage of AI development, we must also remain adaptable.

### **Description of the proposal**


Before AI artists and designers begin creating and grappling with the four checkpoints, we emphasize the need for them to examine why they are using AI tech-




niques in their work in the first place. Prospective users of AI should consider their objectives for implementing AI techniques, how they understand the role of AI technologies in society, and whether or not they are using AI to comment about social or political issues.

*Checkpoint 1: Dataset.* The first checkpoint in the field guide considers the dataset —the foundation of one’s AI work (Figure 3). Artists and designers should think of their training data selection as an inherently subjective act of curation and seek to avoid exploiting other creators’ work or cause harm through what and who is, and is not, represented in the dataset. For example, AI artist Arfa experienced such copyright concerns firsthand when they generated furry persona images from a StyleGAN2 model trained on over 55,000 artworks from the furry fandom, scraped without permission from a furry art forum (Mix, 2020). The original furry art creators protested that Arfa’s project, ‘This Fursona Does not Exist’, disrespected their work, as Arfa benefited from art used without the original creators’ permission or the choice to opt-out. Similarities between

**Figure 3:** Image from the *Responsible IA Art Field Guide* (page 8), by Emily Saltz, Lia Coleman, and Claire R. Leibowicz (Partnership on AI, 2020). Screenshot: Claire R. Leibowicz, 2020. Retrieved from Medium: <https://medium.com/partnership-on-ai/a-field-guide-to-making-ai-art-responsibly-f7f4a5066ee>


**CHECKPOINT 1:  
DATASET** 

The dataset is the foundation of your AI art. Choosing a subset of media as **training data** is an inherently subjective **act of curation**; think carefully about how you select your raw data to avoid exploiting other creators’ work or causing harm through what and who is (and isn’t) represented. Some questions to ask yourself:



**WHERE DOES THE TRAINING DATA COME FROM? WHAT’S MY RELATIONSHIP TO IT?**

- What is the historical and social context of the media I’m using as training data?
- Am I **scraping** data from a public forum or social platform? If so, how do I relate to these communities - in what ways do I have more or less power than other community members?
- Is there content in my dataset which might infringe on a valid copyright, are they in the public domain or creative commons for noncommercial use?
- Am I using an existing dataset? If so, do I understand how and why it was created?

**DATASETS FROM PERSONAL ARTWORKS** 

**Esteban Salgado** is an AI and collage artist who creates his own datasets. Salgado algorithmically generates thousands of abstract vector shapes in Adobe Illustrator, and trains **StyleGAN2 models** on them to create meditative animated blobs.

creators' original works and model outputs also led to complaints of copyright infringement. Beyond copyright, AI artists should also consider the diversity of the dataset and whether or not they are respecting data creators and subjects. In contrast, artist Esteban Salgado creates his own datasets by algorithmically generating thousands of abstract shapes in Adobe Illustrator and training StyleGAN2 models on the shapes (Salgado, 2020). Whether creating one's own data or not, one should consider the historical and social context of the media used as training data, whether or not they are collecting data from public fora or social platforms, and copyright constraints on the dataset.

*Checkpoint 2: Model Code.* Once AI creators decide on a dataset, they must train their models on that data. We encourage anyone considering to use AI for art or design, to learn the history and supply chain of the AI architectures they are using, since doing so can enable them to best respect the people who contributed to their model code, acknowledging the people and labor that went into the code used to produce the work, and critically evaluating how the codebase was developed and labeled. There are complicated ownership questions, too, between AI frameworks, tools, models, and outputs. For example, AI artist Robbie Barrat open-sourced a GAN model that generated fake visuals based on oil painting images. In 2018, artist collective Obvious sold a framed output after duplicating Barrat's neural network method in a piece called 'Edmond de Belamy, from La Famille de Belamy' for 432,500 U.S. Dollars. Barrat received none of this money, raising questions about ownership and credit in the AI art world (Simonite, 2018). Ownership questions are further complicated by anthropomorphized perceptions of AI art as work created by AI as an agent rather than by people using AI as a tool, which was recently explored by Epstein and colleagues (2020).

*Checkpoint 3: Training Resources.* After deciding on data to train and the code to train with, AI makers need a GPU machine(s) and other training resources to actually train models. This process can be very resource-intensive and have a sizeable carbon footprint. Training a single AI model like the popular Transformer deep learning model may emit more than 626,000 pounds of carbon dioxide equivalent to nearly five times the lifetime emissions of the average American car (Hao, 2019). With models commonly trained many times, the emissions for large-scale training can be significant. In order to train models responsibly, artists should consider how they might reduce environmental costs through methods like transfer learning to avoid training models from scratch. Tools like the Machine Learning Emissions Calculator can be used to compute GPU carbon emissions expected from training (Lacoste et al., 2019).

*Checkpoint 4: Publishing & Attribution.* Once artists have trained models and are ready to share their work, we encourage them to be as transparent about their process as possible so that others can learn from their experience.

However, it is also important to remember that others may find and misuse AI work and products for profitable or political motives (Moisejevs, 2019). Thus, artists and designers of AI products should consider the threats and unintended consequences associated with publishing work, and weigh the costs and benefits associated with releasing models, code, and datasets. AI creators can look to the explainable ML field to consider how they make their code accessible to others while limiting the likelihood that motivated malicious actors might seek to weaponize an AI creator's techniques (Bhatt et al., 2020). Several researchers at technology platforms have begun considering ways to publish work to ensure it is not weaponized by malicious actors looking to generate synthetic media to mislead or harm, and artists and designers should do the same (Leibowicz et al., 2020).

---

## CONCLUSION

### Results

The *Responsible AI Art Field Guide* equips artists with emerging best practices and checkpoints to explore in their work. While the bulk of the project sets out to serve as a provocation rather than a prescription, several best practices emerge by the end of the piece. We conclude that the least risky path to take for AI creators is to make their own dataset through original media such as illustration, photography, text, and video. If not, AI creators should credit others' work whenever possible. This goes for makers whose work is in one's dataset, as well as for people who have shared their code. Additionally, if artists and designers are scraping work from the Internet, it is most responsible to prioritize work in the public domain or to directly ask for permission from those whose identity and/or work is featured in the dataset. Creating AI responsibly today also involves paying attention to the environmental impact of one's contributions: AI creators should plan to minimize environmental training resources by using transfer learning from a pre-trained model. Lastly, creators should document their work in detail, to allow others to learn from and critique their process, while remaining sensitive to the potential weaponization of their artistic methods for creating synthetic media with malicious intent. While thorough technical documentation is not typically a practice associated with artistic production, just as it has become more commonplace despite a lack of precedent in the AI developer and research fields, so too should artists using code embed such practices into their workflows (Geburu et al., 2018; Mitchell et al., 2019; Raji & Yang, 2019).

## Evaluation

The *Responsible AI Art Field Guide* project strives to situate the emergent creative AI field in the broader responsible AI and synthetic media communities. This field guide is the first of its kind to center AI artists and designers in responsible AI conversations by considering ways for them to attend to issues in the responsible AI space through referencing concrete case studies. While the field guide currently offers a valuable starting point for AI creators to pursue a “winding path of questioning,” we consider it an initial step and living document for AI creators to reconsider and shape (Saltz et al., 2020, p. 15). We hope that AI creators, like those trained in AI art classes such as the Artificial Images courses, may leverage this guide and provide feedback as they cultivate their skills (Schultz, 2020). Just as many have called for ethics training in computer science classes, so too should art and design classrooms that introduce and teach AI methodologies and tools attend to the ways to think about responsible AI development and deployment (Grosz et al., 2018).

We have begun road testing the usefulness and applicability of the guide with a cohort of 12 graduate students taking the spring 2021 course *Exhibiting Transdisciplinary Research* at the Rhode Island School of Design. The students compiled their own datasets, trained StyleGAN models, and incorporated the generated results into their final exhibitions. Each week they were asked to write reflections on the corresponding checkpoint in the guide: dataset, model code, training resources, and publishing. In addition to compiling their written reflections, we conducted verbal interviews with the students at the end, thus providing valuable feedback on their process of using the guide.

Beyond the classroom, AI artists have begun testing the guide and have documented its use —perhaps as a transparency element for their creative process in and of itself, that underscores their commitment to responsible creative production. The Moving Target Collective, consisting of Alexa Steinbrück, Natalie Sontopski, and Amelie Golfuß, used the guide when producing ‘Latent Riot’, a series of artificial protest signs produced by a generative neural network in 2021. They trained a StyleGAN with images of protest signs from the Boston Women’s March in 2017 that were generated by hand. The artists emphasized how having a distinct guide that synthesized responsible AI principles for the particular challenges of the AI art domain empowered them to create ethically and responsible material.

In the future, we hope other AI artists will treat the guide like a living document and reach out to the creators with feedback and insight from their experiences using the guide. In particular, we are interested in hearing feedback from AI artists that are not only using AI as a tool to create more broadly, but also to interrogate AI as a tool with profound societal impacts. There may be opportunities

for artists to leverage AI tools in artistically evocative ways to contribute explicitly to the responsible AI research field. Understanding how the use of such tools and responsible AI deployment in AI art allows such artists to engage with tricky questions of AI's impact on society is a particularly compelling AI art use case. □

## Conclusion

AI creators can harness the technology's expressive potential, responsibly. Doing so will require reflection and attention to how their work fits into the broader responsible AI field and vast, complex societal dynamics. Derived from both the practical experience of artists and designers using AI and insights of interdisciplinary AI researchers and media experts, the Field Guide is a list of questions and emerging best practices intended as a starting point for AI creators looking to think critically about the societal impact of their work. Doing so can bolster the power of generated imagery to tell stories, beautify, shed light, and even just offer a creative outlet, while mitigating the unintended consequences on information integrity, attribution, rights, and even the environment. □

## REFERENCES

- BHATT, U., ANDRUS, M., WELLER, A., & XIANG, A. (2020). Machine Learning Explainability for External Stakeholders. *Association for Computing Machinery ArXiv*, (arXiv:2007.05408). <http://arxiv.org/abs/2007.05408>
- BICKERT, M. (2020, January 6). Enforcing Against Manipulated Media. *Facebook Blog*. <https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/>
- BUOLAMWINI, J. (2016). *Project Overview Algorithmic Justice League*. MIT Media Lab. <https://www.media.mit.edu/projects/algorithmic-justice-league/overview/>
- CHESNEY, R., & CITRON, D. K. (2018). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107(1753). <https://doi.org/10.15779/Z38RVOD15J>
- COSTANZA-CHOCK, S. (2018). Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice. *Proceedings of the Design Research Society 2018*. <https://doi.org/10.21606/drs.2018.679>
- CRAWFORD, K., & PAGLEN, T. (2019). *Excavating AI: The Politics of Images in Machine Learning Training Sets*. Excavating AI. <https://excavating.ai>
- DIEHM, C., & SINDERS, C. (2020, May 14). "Technically" Responsible: The Essential, Precarious Workforce that Powers A.I. *The New Design Congress Essays*. <https://newdesigncongress.org/en/pub/trk>
- DOLHANSKY, B., BITTON, J., PFLAUM, B., LU, J., HOWES, R., WANG, M., & FERRER, C. C. (2020). The DeepFake Detection Challenge (DFDC) Dataset. *Association for Computing Machinery ArXiv*, (arXiv:2006.07397). <https://arxiv.org/abs/2006.07397v4>
- EPSTEIN, Z., LEVINE, S., RAND, D. G., & RAHWAN, I. (2020). Who Gets Credit for AI-Generated Art? *IScience*, 23(9), 101515. <https://doi.org/10.1016/j.isci.2020.101515>

- GEBRU, T., MORGENSTERN, J., VECCHIONE, B., VAUGHAN, J. W., WALLACH, H., DAUMÉ III, H., & CRAWFORD, K. (2018). Datasheets for Datasets. *Association for Computing Machinery ArXiv*, (arXiv:1803.09010). <https://arxiv.org/abs/1803.09010v1>
- GROSZ, B. J., GRANT, D. G., VREDENBURGH, K., BEHREND, J., HU, L., SIMMONS, A., & WALDO, J. (2018). Embedded Ethics: Integrating Ethics Broadly Across Computer Science Education. *Association for Computing Machinery ArXiv*, (arXiv:1808.05686). <https://arxiv.org/abs/1808.05686>
- HAO, K. (2019, June 6). *Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes*. MIT Technology Review. <https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/>
- HARA, N. (2020). *Pause Fest* [AI-Generated Image]. <http://www.n-hara.com>
- LACOSTE, A., LUCCIONI, A., SCHMIDT, V., & DANDRES, T. (2019). Quantifying the Carbon Emissions of Machine Learning. *Association for Computing Machinery ArXiv*, (arXiv:1910.09700). <https://arxiv.org/abs/1910.09700>
- LEIBOWICZ, C. R. (2020). *The Deepfake Detection Challenge: Insights and Recommendations for AI and Media Integrity*. Partnership on AI. [https://www.partnershiponai.org/wp-content/uploads/2020/03/671004\\_Format-Report-for-PDF\\_031120-1.pdf](https://www.partnershiponai.org/wp-content/uploads/2020/03/671004_Format-Report-for-PDF_031120-1.pdf)
- LEIBOWICZ, C. R., STRAY, J., & SALTZ, E. (2020, July 13). Manipulated Media Detection Requires More Than Tools: Community Insights on What's Needed. *The Partnership on AI*. <https://www.partnershiponai.org/manipulated-media-detection-requires-more-than-tools-community-insights-on-whats-needed/>
- LI, Y., & LYU, S. (2019). De-identification Without Losing Faces. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, 2019*, 83–88. <https://doi.org/10.1145/3335203.3335719>
- LOMAS, N. (2020, August 17). Deepfake Video App Reface is just Getting Started on Shapeshifting Selfie Culture. *TechCrunch*. <https://social.techcrunch.com/2020/08/17/deepfake-video-app-reface-is-just-getting-started-on-shapeshifting-selfie-culture/>
- LYONS, M. J. (2020). Excavating "Excavating AI": The Elephant in the Gallery. *Association for Computing Machinery ArXiv Preprint*, (arXiv:2009.01215). <https://doi.org/10.5281/zenodo.4037538>
- MITCHELL, M., WU, S., ZALDIVAR, A., BARNES, P., VASSERMAN, L., HUTCHINSON, B., SPITZER, E., RAJI, I. D., & GEHRU, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. <https://doi.org/10.1145/3287560.3287596>
- MIX. (2020, May 7). *This AI Spits Out an Infinite Feed of Fake Furry Portraits*. The Next Web. <https://thenextweb.com/news/ai-generated-furry-portraits>
- MOISEJEVS, I. (2019, July 14). Will My Machine Learning System Be Attacked? *Towards Data Science*. <https://towardsdatascience.com/will-my-machine-learning-be-attacked-6295707625d8>
- NICOLAOU, E. (2020, August 27). Chrissy Teigen Swapped Her Face with John Legend's and We Can't Unsee It. *Oprah Daily*. <https://www.oprahdaily.com/entertainment/a33821223/reface-app-how-to-use-deepfake/>
- PARIS, B., & DONOVAN, J. (2019). *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*. Data & Society. <https://datasociety.net/library/deepfakes-and-cheap-fakes/>



- PATRINI, G. (2019, October 7). Mapping the Deepfake Landscape. *Sensity*. <https://sensity.ai/mapping-the-deepfake-landscape/>
- POSTERS. (2019, May 29). *Gallery: "Spectre" Launches (Press Release)*. <http://billposters.ch/spectre-launch/>
- RAJI, I. D., & YANG, J. (2019). ABOUT ML: Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles. *Association for Computing Machinery ArXiv Preprint*, (arXiv:1912.06166v1). <http://arxiv.org/abs/1912.06166>
- RAKOVA, B., YANG, J., CRAMER, H., & CHOWDHURY, R. (2020). Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for shifting Organizational Practices. *Proceedings of the ACM on Human-Computer Interaction, CSCWI*. <https://doi.org/10.1145/3449081>
- ROTH, Y., & ACHUTHAN, A. (2020, February 4). Building Rules in Public: Our Approach to Synthetic & Manipulated Media. *Twitter Blog*. [https://blog.twitter.com/en\\_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media](https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media)
- ROTHKOPF, J. (2020, July 1). Deepfake Technology Enters the Documentary World. *The New York Times*. <https://www.nytimes.com/2020/07/01/movies/deepfakes-documentary-welcome-to-chechnya.html>
- SALGADO, E. (2020, August 5). *Yaku with Circular Loops* [AI-Generated Image]. <https://www.youtube.com/watch?v=ksqw8Q2wv9c>
- SALTZ, E., COLEMAN, L., & LEIBOWICZ, C. R. (2020). *Making AI Art Responsibly: A Field Guide* [Zine]. Partnership on AI. <https://www.partnershiponai.org/wp-content/uploads/2020/09/Partnership-on-AI-AI-Art-Field-Guide.pdf>
- SCHULTZ, D. (2019). *Faces2flowers—Artificial Images*. <https://artificial-images.com/project/faces-to-flowers-machine-learning-portraits/>
- SCHULTZ, D. (2020). *Artificial Images*. <https://artificial-images.com/>
- SIMONITE, T. (2018, November 28). How a Teenager's Code Spawned a \$432,500 Piece of Art. *Wired*. <https://www.wired.com/story/teenagers-code-spawned-dollar-432500-piece-of-art/>
- TWITTER SAFETY [@TWITTERSAFETY]. (2020, October 30). *Our policies are living documents. We're willing to update and adjust them when we encounter new scenarios or receive important...* [Tweet]. Twitter. <https://twitter.com/TwitterSafety/status/1322298208236830720>