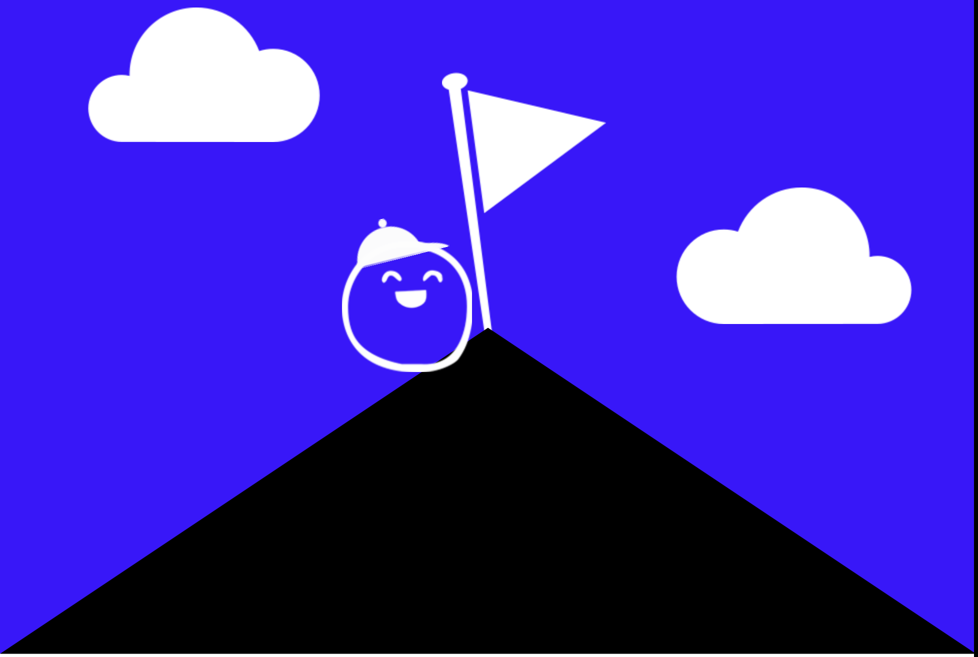
MAKING AI ART RESPONSIBLY

A FIELD GUIDE



PARTNERSHIP ON AI

EMILY SALTZ LIA COLEMAN CLAIRE LEIBOWICZ

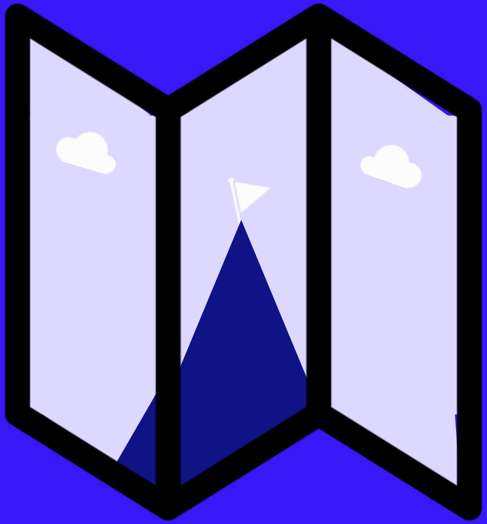
THE RESPONSIBLE AI ART FIELD GUIDE

WHAT IS THIS GUIDE, AND WHO IS IT FOR?

This guide is part of the Partnership on AI's AI and Media Integrity program that investigates emerging AI technology's impact on digital media and online information.

It is a practical field guide to help artists and makers create art using AI techniques responsibly and with care.

Written by Emily Saltz, Lia Coleman, and Claire Leibowicz and illustrated by Lia Coleman. Inspired by a July 2020 presentation with Gray Area, a cultural hub for art and technology based in San Francisco, on "How to Use AI for your Art Responsibly."



HOW TO USE THIS GUIDE

We structured this guide around questions to ask yourself while making AI art. We don't always provide clear answers, because when it comes to the nascent and evolving AI art field, many topics are still subject to ongoing debate. We leave you with some emerging best practices and checkpoints to try out in your work.

NOTE

If you'd like to try it on your own or as part of a group or class, let us know how we can help you document the experience and refine this guide by contacting aimedia@partnershiponai.org.

FIRST, SOME DEFINITIONS:

WHAT IS AI ART?

There are many definitions of artificial intelligence (AI), but here's how we're defining AI art:

AI Art

=

New works made with creative intent

using techniques where computer programs access data and use that data to learn for themselves, such as machine learning.

In this guide we focus on machine learning for image, video, and text generation.

WHAT IS RESPONSIBLE AI?

Responsible AI = The practice of developing and ensuring AI technologies are ethical, transparent, and inclusive for the benefit of people and society

For definitions of more terms in this guide, flip to the glossary in the back.

HI!
I'M AN ARTIST WHO WANTS TO USE AI IN MY WORK!

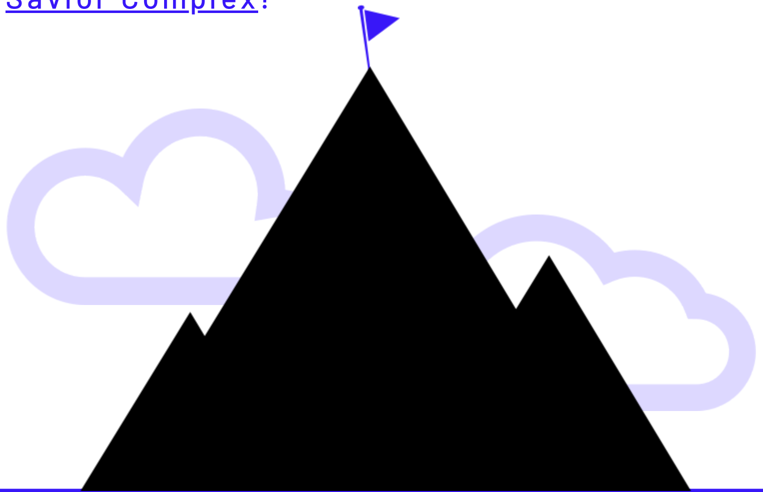
YOU'LL NEED:

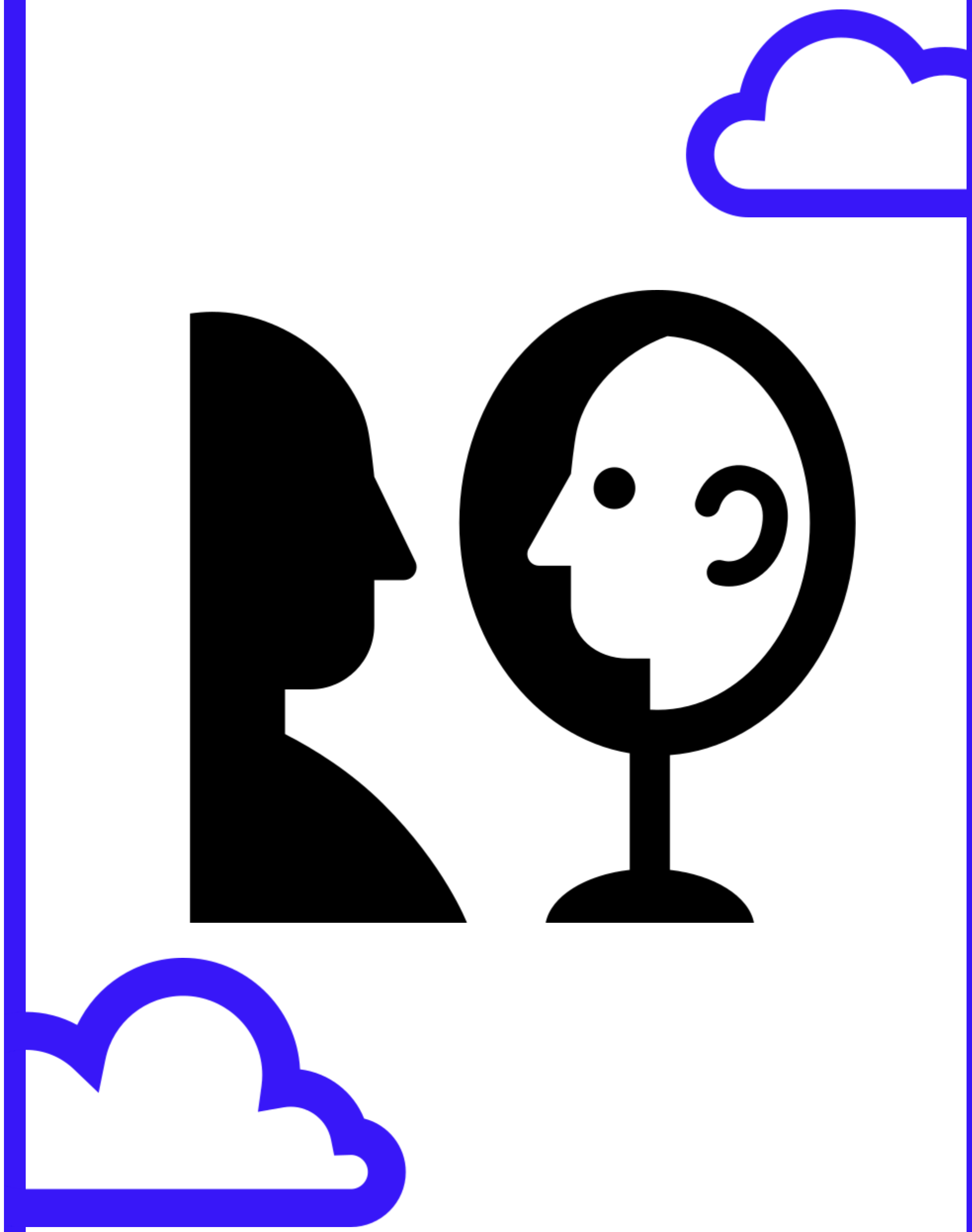
- DATASET
- MODEL CODE
- TRAINING RESOURCES
- A PUBLISHING PLAN

HOLD UP! WHY AM I EVEN MAKING AI ART?

Before we begin our ascent, check in and ask yourself why you're using AI techniques in the first place:

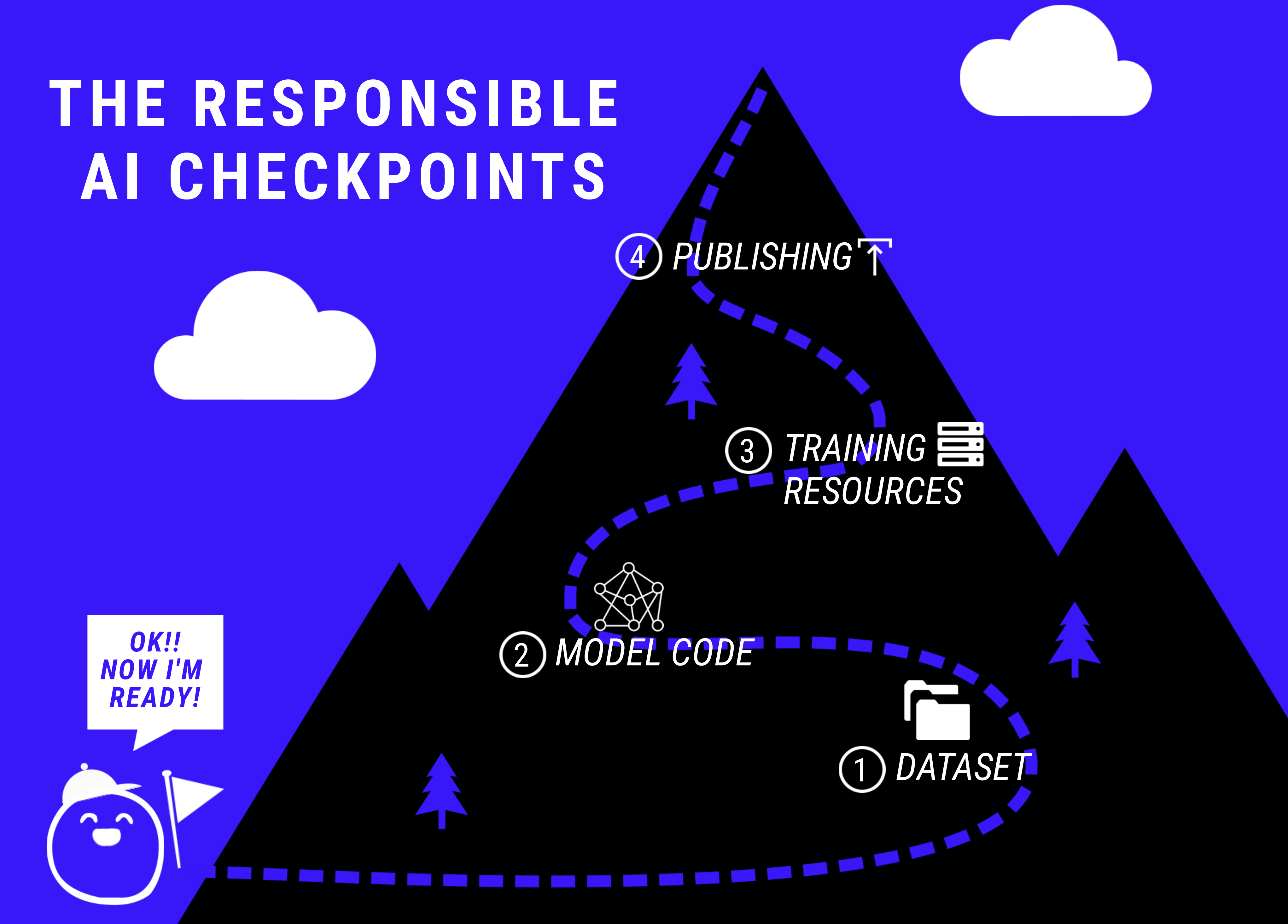
- □ What are my objectives for using AI in this work?
- □ Do I plan to get paid for this AI art? Am I seeking fame? Notoriety? Money?
- □ What are the pros and cons of using AI for this work? Can my objectives be accomplished any other way?
- □ How do I understand the role of AI technologies in society? What might it mean to create Computer Critical Computer Art?
- □ Am I using AI to comment about social or political issues? If so, am I suffering from a Creative Savior Complex?





Be honest with yourself about your goals for using AI, even if you're just looking to learn and play!

THE RESPONSIBLE AI CHECKPOINTS



1 DATASET

2 MODEL CODE

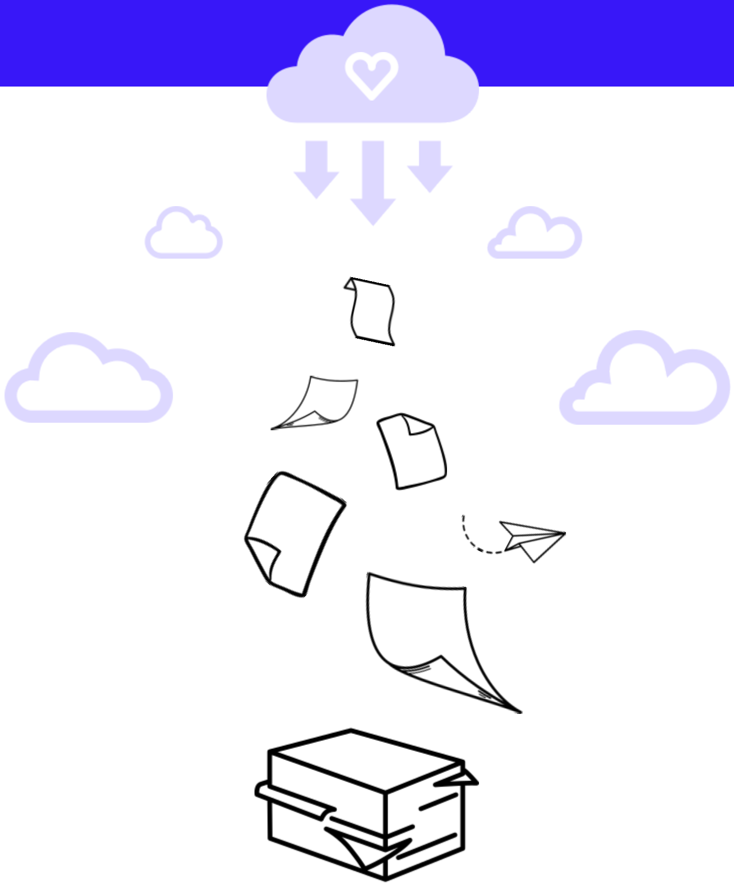
3 TRAINING RESOURCES

4 PUBLISHING

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.

CHECKPOINT 1:

DATASET

HOW DIVERSE IS THE DATASET?

- □ What is represented in the dataset? What data might it be missing, and why?
- □ Are there ways that my dataset might be skewed in a way that would produce ***outputs*** that reinforce stereotypes about race, gender, or other traits?
- □ Is the training data diverse enough to ensure the model doesn't produce near-copies of the original data?

SCRAPING PUBLIC FORUMS

“This Furson Does not Exist” by Arfa generates furry persona images from a StyleGAN2 model trained on over 55,000 artworks scraped without permission from a furry art forum.

Creators of the original furry art protested that the project disrespected their work, as Arfa was benefiting from art used without permission or the choice to opt out. Similarities between creators' original works and model outputs also led to complaints of copyright infringement.

DATASETS FROM ORAL HISTORIES

Stephanie Dinkins is a transmedia artist and professor at Stony Brook University who creates custom AI systems from small, community-focused data, especially with communities of color.

In **“Not the Only One,”** Dinkins creates a custom voice-interactive AI “memoir” trained on interviews with three generations of a Black family in America.

AM I RESPECTING DATA CREATORS AND SUBJECTS?

- □ How might I contact the creators of the data for permission to use their work?
- □ How might I collaborate with the people who created the original data and include them in my process?
- □ How might I collaborate with the people featured in my dataset and include them in my process?
- □ Do the people featured in my dataset even know they're in my dataset?

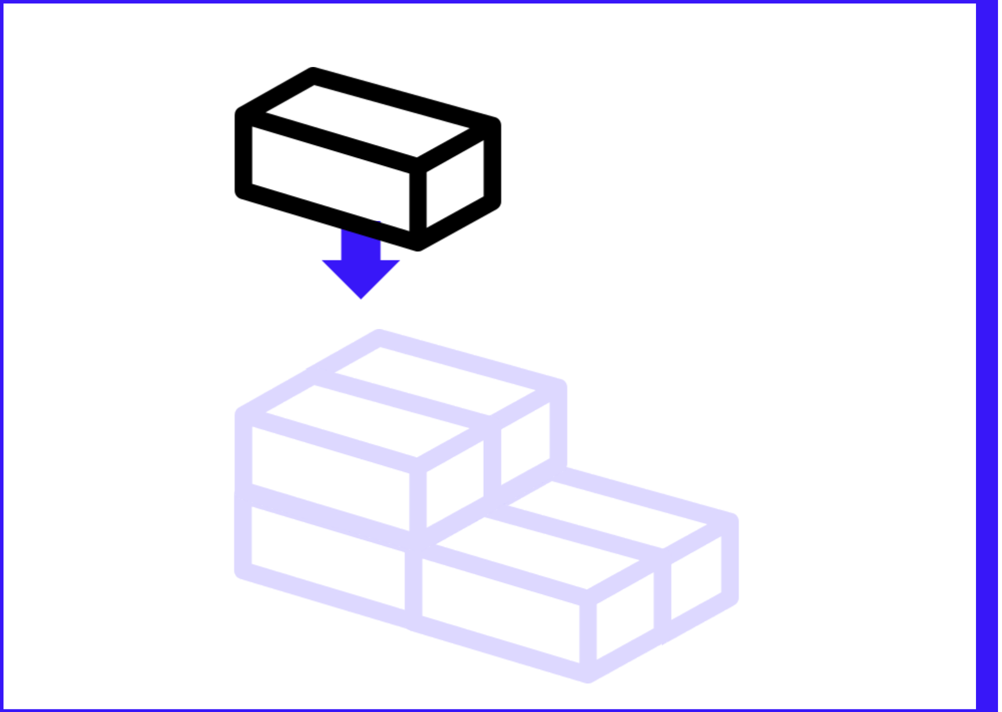
CHECKPOINT 2:

MODEL CODE

Now that your dataset's ready, it's time to make a plan for ***training***.

Most likely, your ***model code*** will be building off existing frameworks developed by researchers at universities, government agencies, or companies like Nvidia with goals ranging from scientific research to military intelligence. Learn the history and supply chain of the AI architectures you're using – this too is part of your work.

Ask yourself:



WHOSE CODE ARE YOU DEPENDING ON TO MAKE YOUR WORK?

- □ What is my relationship to the creators of the tools I'm using?
- □ Am I comfortable with that?
- □ How was this codebase developed and labeled, and by whom?

THE POLITICS OF CLASSIFICATION

The code you're using may also depend on datasets that have been annotated by people on platforms like Amazon's Mechanical Turk for low wages. In **Excavating AI**, Kate Crawford and Trevor Paglen detail examples of "problematic, offensive, and bizarre" labels on ImageNet, a large visual database designed for use in visual object recognition software.

CHECKPOINT 2:

MODEL CODE

AM I RESPECTING THE PEOPLE WHO CONTRIBUTED TO THE MODEL CODE?

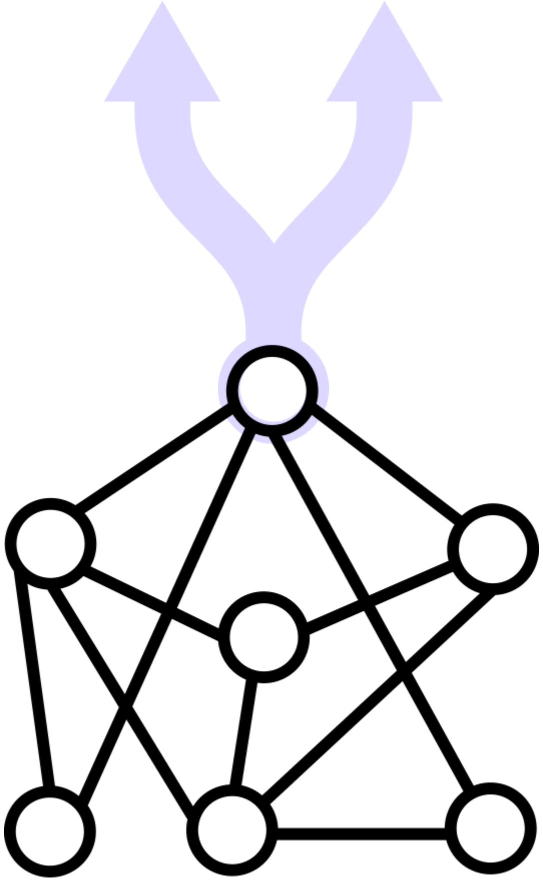
- □ If I'm using someone else's code, am I thanking and crediting them?
- □ Am I appropriately acknowledging the people and labor that went into the code used to produce the work?

WHO OWNS AI ART? AI MODELS & MODEL OUTPUTS

Robbie Barrat is an AI artist who open-sourced a ***GAN*** model that generated fake visuals based on images of oil paintings.

In 2018, artist collective Obvious **sold a framed output from Barrat's neural network** in a piece called "Edmond de Belamy, from La Famille de Belamy" for $432,500. Barrat received none of this money.

This example raises the question — still legally ambiguous — about ownership relationships between AI frameworks, tools, models, and outputs.

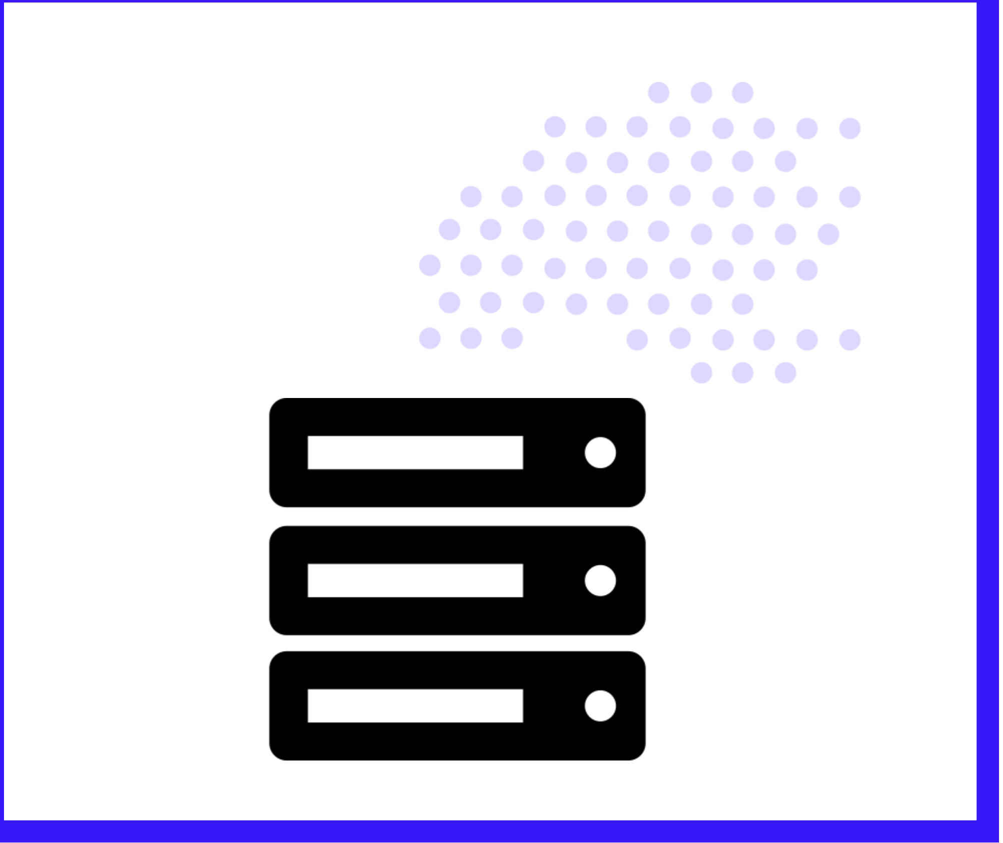


CHECKPOINT 3:

TRAINING RESOURCES

Now that you have data to train and the code to train with, you'll need a ***GPU*** machine(s) and other ***training resources*** to actually train your model.

This training can be very resource intensive, with a sizable carbon footprint: to check yourself, compute GPU carbon emissions expected from training through tools like this Machine Learning Emissions Calculator. Ask:



WHAT ARE THE ENVIRONMENTAL COSTS OF MY TRAINING?

- □ How might I reduce environmental costs through methods like ***transfer learning*** to avoid training from scratch?
- □ Do I need to train a new model at all? Can I use existing pre-trained models?
- □ How long do I need to train to reach outputs I'm satisfied with? How can I ensure I don't overtrain with diminishing returns?

TRAINING A MODEL THAT EMITS THE CARBON OF FIVE CARS

Training a single AI model like the popular **Transformer deep learning model** may emit "more than 626,000 pounds of carbon dioxide equivalent—nearly five times the lifetime emissions of the average American car (and that includes manufacture of the car itself)," as reported by MIT Tech Review.

Considering that these models are typically trained many times and by many researchers, the emissions add up.

CHECKPOINT 4:

PUBLISHING & ATTRIBUTION

Finally, you’ve trained a model and have something you’re ready to share with the world! Make sure you share everything else you’ve learned while scaling this mountain: the more you share about your process, the more others can learn.

At the same time, just because you intend only to create art for an artist community, doesn’t mean others can’t find and misuse your work for profit or political motives – it’s crucial to consider the threats and unintended consequences associated with publishing your work. Ask:

WHO MIGHT BENEFIT FROM THIS WORK?

- □ Will this be in a show? Will I get $$ or publicity from this?
- □ How am I crediting and thanking the people involved with the model code and dataset?

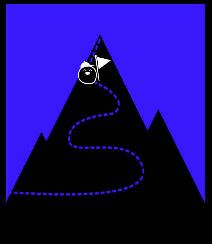
CREDITING INDIVIDUAL FORUM CONTRIBUTORS

In Everest Pipkin’s **i've never picked a protected flower (concrete unicode poems)** book of poems, they use titles scraped from forum posts of users of wordreference.com, an ESL forum for translating phrases and idioms. While not AI-generated, Pipkin's scraping technique is similar to the dataset creation process used for training models.

Pipkin credits hundreds of names of forum users in the back of the book, acknowledging the profound way that this work depended on the forum users’ work. Additionally, they open-sourced the code used to generate the poems, and provide an opportunity for contributors to opt out.

CHECKPOINT 4:

PUBLISHING & ATTRIBUTION



WHAT ARE UNINTENDED CONSEQUENCES OF RELEASING MY MODEL / CODE / DATASET?

□ Have I thought through who might be motivated to use and misuse my work, how, and why, e.g. for political or profit motives?

□ Does my model/code/dataset contain sensitive or confidential information that could be used to identify individuals, such as addresses, information about health, or political or religious beliefs?

□ If so, can I password-protect the work, or release to a closed community instead?

□ Will I be storing this model/code/dataset in perpetuity? If not, how long will I store them?

□ Have I checked with people involved in my dataset creation and model code? How might my decision to release or not release this work affect them?

RECONSTRUCTING PRIVATE TRAINING DATA

For technically sophisticated actors, e.g. those undertaking state-sponsored attacks, even private features of machine learning models may be recovered, as described in **Microsoft security documentation** about techniques for reconstructing private training data.

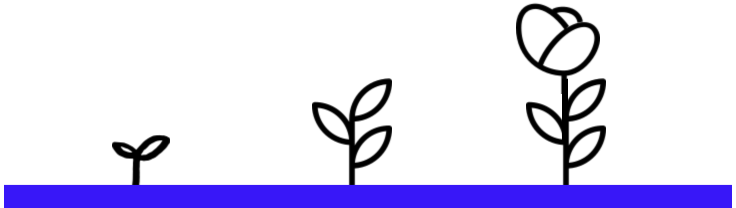
TRANSPARENT AND COMPREHENSIBLE AI DOCUMENTATION

The Partnership on AI has several projects developing with recommendations for how AI researchers can better document how their AI technologies were created and why – including **ABOUT ML**, **Publication Norms**, and **Explainable ML**.

HOW MIGHT I MAKE MY WORK ACCESSIBLE TO OTHERS TO SUPPORT DISCUSSION AND LEARNING?

□ Have I documented my work with explainable AI fields to be transparent about the work that went into this project?

□ Are outputs accessible to people with varying access needs using video and image descriptions?



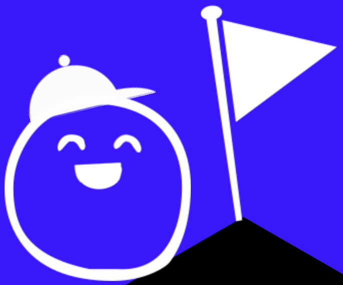
BEST PRACTICES

Least risky:
Make your own dataset through ***original media*** such as illustration, photography, text, and video.

If you are scraping work, prioritize work that is in the public domain,

or directly ask for permission from those whose identity and/or work is represented in the dataset.

Save environmental training resources by ***transfer learning*** from a pre-trained model.



Credit the work of others whenever possible!

If you are posting online, tag people and thank them.

This goes for artists whose work is in your dataset, as well as people who have shared their code or

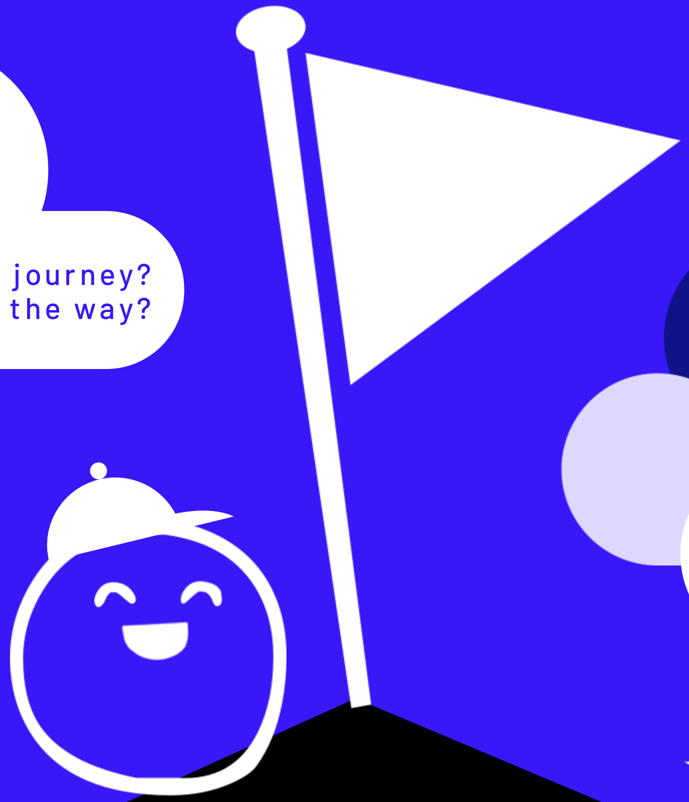
Document your work in detail to allow others to learn from and critique your process, and access your work.

YOU'VE CLIMBED THROUGH A WINDING PATH OF QUESTIONING TO REACH THE FINAL CHECKPOINT: THE BEST PRACTICES SUMMIT.

YOU DID IT!

So you made it past the checkpoints, and have ascended, ready to share your AI art from the mountaintops.

How was your journey? Any stumbles along the way?



Reflect on your process and share what did and didn't work with this guide

by contacting aimedia @partnershiponai.org

HAPPY AI ART-MAKING!

Emily, Lia, and Claire

RESOURCE & REFERENCE LIBRARY

Take a break from AI art-making to browse the resource library and learn more about the works that went into these checkpoints.

DATASET

Missing Datasets, Mimi Onuoha

Case Study: Esteban Salgado

Case Study: This Furson Does not Exist, Arfa

Case Study: Not the Only One
Stephanie Dinkins

Corpora as medium: on the work of curating a poetic textual dataset, Everest Pipkin

TRAINING RESOURCES

Machine Learning Emissions Calculator
ML CO2 Impact

Training a single AI model can emit as much carbon as five cars in their lifetimes
MIT Technology Review

PUBLISHING PLAN

Threat Modeling AI/ML Systems and Dependencies - Security documentation, Microsoft

Will my Machine Learning System be attacked? | by Ilja Moisejevs

ABOUT ML, PAI

Publication Norms, PAI

Multistakeholder Approaches to Explainable Machine Learning, PAI

Alternative Text, WebAim

WHAT IS RESPONSIBLE AI

How to Use AI for your Art Responsibly
Our Gray Area Webinar that serves as the basis for this guide

Computer Critical Computer Art, Sarah Groff Hennigh-Palermo

How to think differently about doing good as a creative person
Omayeli Arenyeka

MODEL CODE

History of artificial intelligence, Wikipedia

Case Study: Excavating AI
Kate Crawford and Trevor Paglen

ImageNet

MegaPixels
Adam Harvey and Jules LaPlace

Case Study: "Edmond de Belamy, from La Famille de Belamy",
Robbie Barrat and Obvious, via Wired

LIFELONG LEARNING ROOM

Practical Data Ethics, fast.ai

Tutorial on Fairness Accountability Transparency and Ethics in Computer Vision at CVPR 2020, Timnit Gebru and Emily Denton

Artificial Images, Classes by Derrick Schultz & Lia Coleman

Gray Area, San Francisco, CA

The School for Poetic Computation, NYC, NY

GLOSSARY

DATASET

Training data: Any collection of data used as a basis for training a machine learning model

Scraping: Techniques for extracting media, such as text and image, from documents in order to create a dataset

Training: The process, lasting from hours to days, where a ML algorithm learns how to create new media based on patterns in the training data

MODEL CODE

Model code: The code containing AI algorithms which are used for training, often accessed through Github repositories from paper releases, or technology companies like Nvidia

Model: The artifact created at the end of the training process which is used to generate novel works based on the training data

Model output: An instance of novel media generated by the AI model based on the training data media

GAN: Generative Adversarial Network. Machine learning frameworks in which two neural networks, a generator and a detector, compete with each other, leading the generator to learn statistical patterns from the training data and generate realistic new media

StyleGAN2: A GAN framework in use by many AI artists, created by Nvidia researchers and released to the public in February 2019

Transfer learning: A technique used for more efficient training in which a pre-trained model is used as the basis for training a new model based on a different dataset

TRAINING RESOURCES

Training resources: The computational machinery used for training AI algorithms, such as Nvidia's CUDA software, GPUs and TensorFlow. These resources can be accessed either through tools providing GPU infrastructure like Paperspace, CoLab, GCP, or Runway, or else using your own GPU

GPU: The graphics processing unit of a computer alters computer memory to create new images for display

ABOUT THIS GUIDE

WHO MADE THIS GUIDE AND WHY?

This guide is part of the Partnership on AI's AI and Media Integrity program that investigates emerging AI technology's impact on digital media and online information.

The guide showcases case studies and processes we've learned through PAI's broader work on synthetic and manipulated media, machine learning explainability, FTA (fairness, transparency, and accountability), and publication norms for responsible AI. This guide is also informed by Lia Coleman's personal experience with the common challenges faced as an AI artist and AI arts educator. We hope you'll reference these questions as a starting point as you use AI methods to generate your own creative AI works responsibly.

I'M AN ARTIST, WHY SHOULD I CARE?

As artists and other independent creators experimenting with AI technologies, it's crucial to recognize that as you create AI art, you are also a de facto AI researcher. By releasing AI art into the world, you are responsible for understanding the potentially harmful unintended consequences of your work.

Beyond this, without institutional guardrails, you have the opportunity to be scrappier and more creative in achieving your goals – often breaking and pushing the state of the art beyond the traditional, narrow use cases of machine learning research largely funded by state and corporate stakeholders.

The state of the art in machine learning changes fast, and use cases for art and accessibility can be key drivers of innovation. You wield power to define AI practices by taking care with how you create work to serve as a model for others in the AI research space.

ACKNOWLEDGEMENTS

Thanks to feedback from the Partnership on AI staff and Partners, Artificial Images and School for Poetic Computation communities, as well as attendees of Gray Area's “How to Use AI for your Art Responsibly” webinar in July 2020.

https://www.partnershiponai.org/
115 Sansome St. Ste. 1200
San Francisco, CA 94104