

---

Pereira has a PhD degree from Aarhus University (Denmark). His research focuses on critical studies of data, algorithms, and digital infrastructures, particularly those of computer vision. <https://www.gabrielpereira.net/>

**Bruno Moreschi** is the co-coordinator of the Group on Artificial Intelligence and Art (GAIA)/C4AI at the University of São Paulo. He is a researcher of the project Decay without mourning: Future thinking heritage practices, Riksbankens Jubileumsfond/Volkswagen Foundation. Moreschi received his PhD in Arts at the State University of Campinas (Unicamp). Moreschi's investigations are related to the deconstruction of systems and the decoding of social practices in the fields of arts, museums, visual culture and technologies. He is a Senior Research Fellow at the Center for Arts, Design + Social Research.

---

## Appendix

### Attachment 1: Teaching guide for “Living with images from large-scale data sets”

*(This exercise guide is just a skeleton. Please adapt it to your students' needs and possibilities, and complement it with discussion and exposition that fits your class' goals and framing.)*

**Learning goals:** 1) Students will apply a practice-based research method to explore the formation of data sets through visual elements; 2) Students will analyze and reflect on their experience with these images, and discuss them in-class as they relate to issues of data set creation and use; 3) Through building responses to images and in-class discussions, students will build critical knowledge on issues around the scale of data sets.

**Prior knowledge required:** This exercise is well-suited for students at any level, particularly in courses dealing with critical approaches to data and algorithms in society. It fits well with other activities around the curation and use of data sets. If students are not at all familiar with data/algorithms, a brief explanation of the operation of computer vision is suggested.

**Time required:** This plan includes a take-home exercise. In the first class, students briefly discuss issues around data sets and their scale, as well as receive images to live with. Between the first and second class, at least one week passes where students live with their images and prepare any form of response. In the second class, students are asked to share their responses and discuss them. The second class concludes with the takeaway points from the exercise, as they relate to the development and use of data sets in society.

**Materials required:** 1 to 3 printed images per student. If it is not possible to print images, give each student a PDF file or ask them to pick their own images. See Attachment 2 for a tool that automates the process of generating images to live with. Please always double-check all images before giving them to students to avoid issues with sensitive images.

**Suggested readings:** Denton et al, “On the genealogy of machine learning datasets”; Malevè, “On the data set's ruins”; Thylstrup, “The ethics and politics of data sets”; Crawford and Paglen, “Excavating AI”; Prabhu and Birhane, “Large image datasets”; and Smits and Wevers, “The agency of computer vision models”.

**First class: Introduce the exercise.** (Around 30 minutes.)

- Ask students: Have you ever seen a data set? Do you think your images are part of a data set? What do you think happens when images are added to data sets?
- Use students' responses to set up for the assignment. You can engage students by showing the grandeur of data sets millions of images and/or with real-life cases and the questions that emerge from them, e.g.:
  - Clearview AI and their construction of a database indexing 20 billion images of people's faces, without their consent.
  - The case of ImageNet Roulette and the consequences of this intervention.
- Set up *why* students will be doing this assignment:
  - Some would call these “images that should not be seen”, as their goal is just the support of machine learning models. These are large-scale databases, with millions and millions of items. But what could happen if we engaged with them from up

close? What could be learned by looking at the images in themselves, not as a large-scale collection?

- Give students three images each.
- Ask students to write down what they see in this image in 20 seconds. Explain this is their initial impressions, which is what computer vision algorithms most often focus on (the tags created by microworkers). Our goal, however, is to get closer to such images and see what else is in them that we might not see at first glance. (Ask students to keep this first description for next class.)
- Explain the assignment:
  - You will be taking home three images. You should live with these images in whichever way you feel comfortable, and prepare a response. Try to create reflections on this process as you go, not only in the very end. These can take shape in any (multimedia) format: images, drawings, videos, etc.
  - Complete instructions will be published on student LMS (see section below).
- Set students' expectations that living with some of these images can be hard because they are pixelated and often may seem boring at first glance. If this difficulty comes up, treat it as part of the exercise and reflect on it!

#### **Assignment instructions to be published in student LMS**

This assignment asks that you live with three images from large-scale computer vision data sets. You have received (or will receive) three images from your instructor.

To live with an image means to have it around you, to spend time with these images by looking at them at least for a little bit of time every day. To remember to do this, you may want to place it somewhere they'll be seen, for example: on the table where you have breakfast, next to your computer screen, or even next to your toothbrush. You may also want to take your images on a hike, to get them out there to experience the world! Make it fun!

As you live with your image, you may want to take notes of your thoughts or any things that come up. You can write down your thoughts on your phone, computer, notepad, or even the back of the image or on the image itself.

Everything matters in getting closer to this image, so try to write things up! Here are some questions to get you going, but many others are possible:

- What do you think is the story of your image(s)? How would you tell them to your friends and family? Does this story change as you look more and more at this image?
- How do you think your three images relate? What images from your life would you like to put next to them? What relations emerge from the connection between these images and new images?
- How do you think these images have been used before? Where do you think they come from, and how do you think they are to be used by algorithms? How do you think an algorithm would understand it, and why?

For the next class, please come with your responses from living with your images. Your response can be the notes you've been taking, but if you feel creative, you can also bring something else. For example, you may write a poem about your image, create other images based on them, compose a song, or even film something with your phone. There are no limits to your creativity, as long as it relates to and reflects on your experience of these images.

*(Depending on your preference, you may request students to publish their responses in the LMS before the second class.)*

#### **Second class: Post-assignment discussion.** (From one to two hours.)

- Start the class with a prompt: In 30 seconds write down what you see in this image now that you've lived with it. How is it different than your initial perception?

- Open up the class for students to discuss their experiences and invite them to share their responses, e.g.:
  - Where did you place your images? What did it mean to live with them for you? What unexpected things happened by living with them?
  - How do you think getting close to them changed your experience of the images? Get them to show the images and share their responses with the class. (This can be done in small groups if class size allows.)
- Present key points around data sets and scale, by linking the experience of students to the suggested readings and/or exposition by the lecturer, e.g.:
  - Data sets are formed around simple descriptions or tags that are given to images. These descriptions are limiting and can be problematic.
  - There is an assumption of scale across the production of computer vision models. This leads to many potential problems, including in the provenance of data, microworker labor, and the right to deletion.
- Discuss take-away points with students, e.g.: How should data sets be developed and used? The project 'Data sheets for data sets' sought to engage data set developers in documenting data sets, including potential future use and how to maintain the data.

#### Attachment 2: 'Images to live with' tool

We have created an online tool to automatically generate PDFs containing images from large-scale image data sets.

The tool is available at

<https://learningfromlivingwithdata.herokuapp.com/>

or at

<https://www.gabrielpereira.net/imagestolivewith/>

#### **Images to live with**

This is a tool to create PDFs with data set images for printing.

To generate a PDF with images to live with, fill the form below with the number of images you want (max 50) and press OK.

The tool automatically picks from a selection of 14436 images randomly chosen from 8 different datasets: ("Tiny ImageNet", "COIL-100", "Labelled Faces in the Wild", "Leaf", "Open Image", "Stanford Dogs", "Tencent", and "VisualGenome").

For more details, read the article "Living with images from large-scale data sets: A critical pedagogy for scaling down", on the Photographies journal.

Full code can be found on [GitHub](#).



Created by [Gabriel Pereira](#) and [Bruno Moreschi](#).