<PREDECENT STUDY id="ImageNet Roulette | Excavating AI" >

<head><title> Image Datasets | Training AI | Digital Provencance </title>

<meta name="author" content="Varun SA" /></head>

<body>

<Image Datasets id="ML Training">

 Image datasets are curated collections of labeled visual content that serve as foundational material for training artificial intelligence (AI) models, particularly in the field of computer vision. As of now, these datasets are essential for enabling AI systems to learn how to interpret and make sense of visual information --- whether the task is to identify objects, detect features or even generate entirely new images.

 Traditional image datasets ---> 'ImageNet', 'COCO', 'MNIST', 'CIFAR-100' ---> are used to train Discriminative models ---> models which recognize and classify images based through training on example images and their labelled text.

 Generative AI models (like Dall.E, Stable Diffusion, MidJourney) ---> are trained on datasets that pair images with text descriptions ---> such as 'LAION-5B', 'Conceptual Captions'.

 Image datasets are made by ---> 1) Data collection ---> 2) Data cleaning & filtering ---> 3) Labelling & Captioning.

<li class="Types of Labels"> Labels are used to correlate information with the images or features in the image. They are also used to group, classify and arrange images. They include: 1) Class labels, 2) Bounding boxes, 3) Segmentation masks, 4) Captions, 5) Alt-text, 6) Keypoints.

There is no neutral or apolitical foundation on which image training datasets can be built. The very process of collecting, labeling, and categorizing images is political—raising questions about who decides what images mean and how they are used.
 ImageNet is a striking example, having once included categories such as "failure," "loser", "drug addict," "slut," and even racial slurs and misogynistic terms.
 The recognition and interpretation of an image are always subjective, varying across social, cultural, and historical contexts. Trevor Paglen's ImageNet Roulette demonstrated the dangers of training AI on such flawed data: when these biased categories are operationalized by algorithms, they can reproduce harmful stereotypes at scale.
 Far from being purely technical, automated image interpretation is a deeply social and political project, yet it is increasingly embedded into institutions where it can influence or even make decisions.

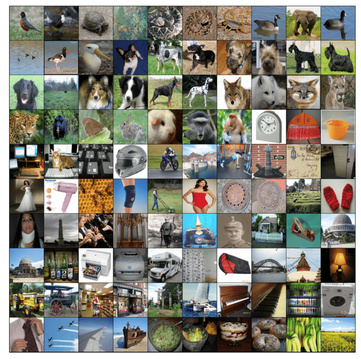
For decades, the ability to interpret images was considered a hallmark of intelligence, and today object detection and facial recognition are seen as largely "solved" problems. But these systems rest on a fundamental limitation: images are flattened into labels that assume fixed, universal meanings, with little or no context.
 In ImageNet, categories are organized around WordNet's taxonomy of nouns—ranging from concrete to abstract—forcing complex, ambiguous concepts into simplistic labels.
 This reduction strips away the social and historical layers that shape meaning, and when applied to people, it can reproduce pseudoscientific ideas reminiscent of physiognomy or eugenics, correlating features like skin tone or facial expression with moral or criminal traits.
 These datasets assume a direct correspondence between image and concept, but this is not how the world works. As a result, AI trained in this way lacks deep understanding, and while it may succeed at narrow tasks like object recognition, it fails dramatically when applied to complex phenomena such as crime detection or behavior analysis.

Most large datasets are built by scraping images from the internet without people's knowledge or consent. Individuals lose ownership of their data, receive no compensation, and often see their images used in technologies like surveillance systems that may harm them. Even when problematic datasets are taken down, their influence lingers because trained models and copies continue to circulate, highlighting the exploitative nature of current AI data practices.

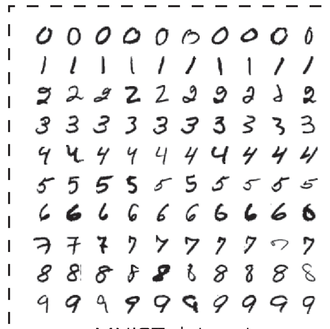
</Image Datasets>

</body>

</PREDECENT STUDY>

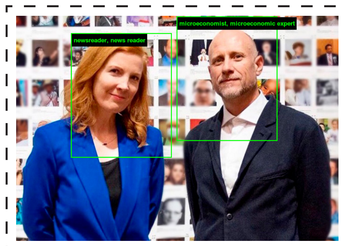


ImageNet sample photos
Source of images: Web image search (Google, Flickr etc.)



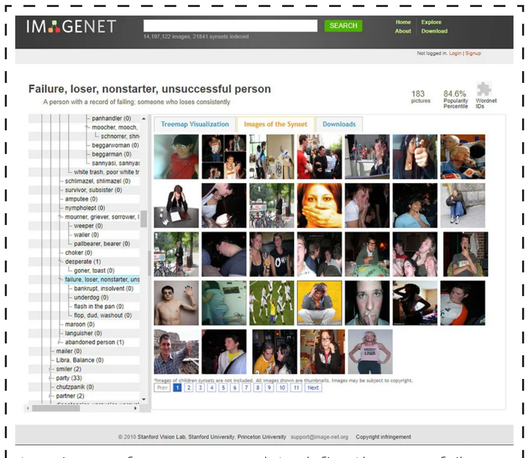
MNIST dataset
Images of handwritten digits from U.S. census bureau

COCO - Common Objects in Context
Segmentation mask labelling

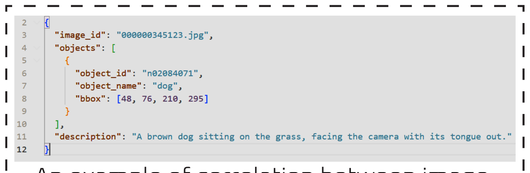


Kate Crawford and Trevor Paglen use the ImageNet Roulette
Bounding-box labels

"This is not an Apple"
by Rene Magritte
is a surrealist painting that questions the relationship between images, words, and reality. Although it shows a realistic apple, the inscription reminds us that it is only a representation, not the actual object. The piece highlights how signs—whether visual or linguistic—are distinct from the things they refer to. By disrupting the automatic link we assume between image, word, and object, Magritte encourages viewers to reflect on perception, meaning, and how easily representation can mislead or oversimplify reality.

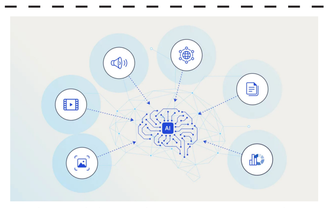


Is an image of a person enough to define them as a failure, a loser, or an unsuccessful individual?
Can someone be labeled a nonstarter simply because of their appearance, clothing, or facial expression?
Are images and labels alone sufficient to understand reality?



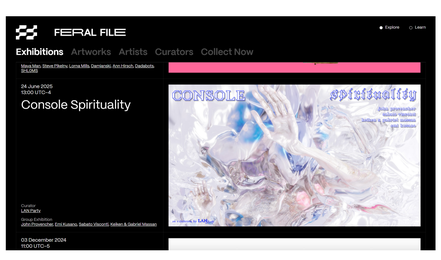
```
{
  "image_id": "000000345123.jpg",
  "objects": [
    {
      "object_id": "n02084071",
      "object_name": "dog",
      "bbox": [48, 76, 210, 295]
    }
  ],
  "description": "A brown dog sitting on the grass, facing the camera with its tongue out."
}
```

An example of correlation between image, object and a description.



Multimodal simulation AI can combine information from text, images, sound, 3D models, and dynamic environments, giving it a richer, more contextual understanding of the world than bimodal systems. By learning from how different modes interact, it can capture nuance, causality, and real-world complexity.

Feral File
is a digital art platform that uses blockchain to establish digital provenance, presenting artworks as NFTs and ensuring artists are recognized and compensated for their work.



Can blockchain technology evolve beyond its current use in digital art ownership to safeguard all digital assets—from everyday data like text and photos to highly complex files such as 3D digital twin models used in simulation?