Casey James

INFM480

## How Generative Image AI Works

In order to better understand the subject matter, a simplified overview of the dynamics and functionality of generative image AI is provides.

What is generative image AI?

Generative image AI refers to Artificial Intelligence models that are able to create imagery through text or image prompts. These AI models are trained using a dataset, in this case billions of images, that is curated for the use of the model (*Research guides: Artificial intelligence for image research, n.d.*)

According to Medium, (Dominguez, 2024) they can be broadly categorized into two categories: Variation Autoencoders (VAEs) and Generative Adversarial Networks (GANs).

- Variational Autoencoders (VAEs): VAEs are probabilistic models that encode images into a latent space, where they are represented as vectors. The decoder then reconstructs the images from the encoded vectors, enabling the model to generate new images by sampling from the latent space.
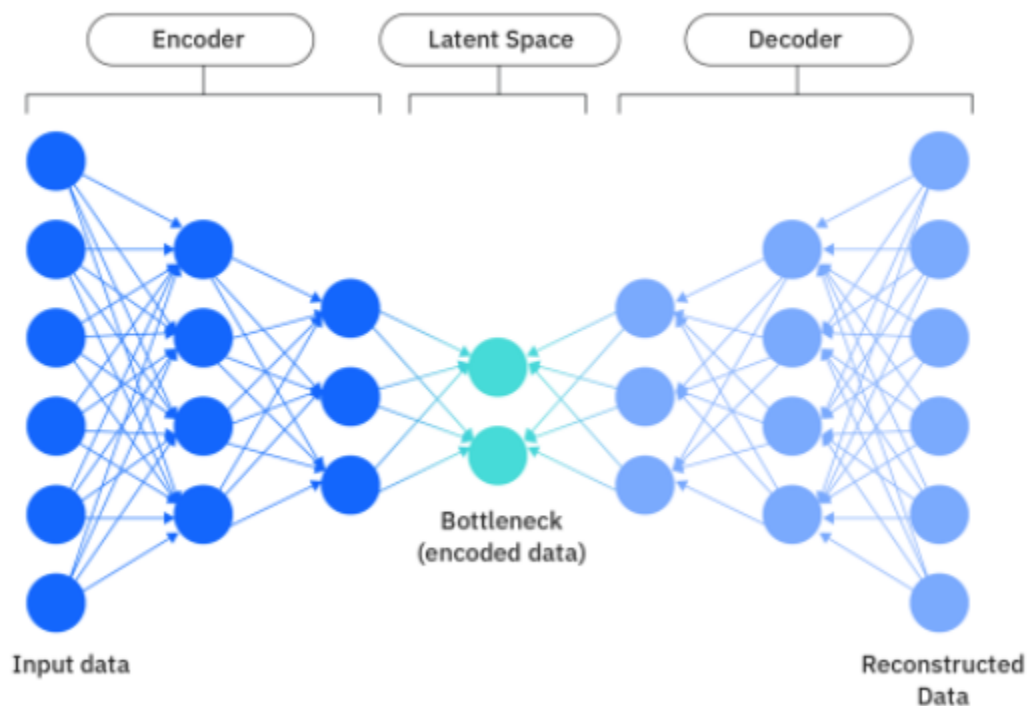
(Bergmann & Stryker, 2024)

- Generative Adversarial Networks (GANs): GANs consist of two neural networks, a generator, and a discriminator, engaged in a competitive process. The generator creates synthetic images to fool the discriminator, which, in turn, aims to distinguish between real and fake images. This back-and-forth battle results in the generation of highly realistic images.

GENERATIVE ADVERSARIAL NETWORK ARCHITECTURE
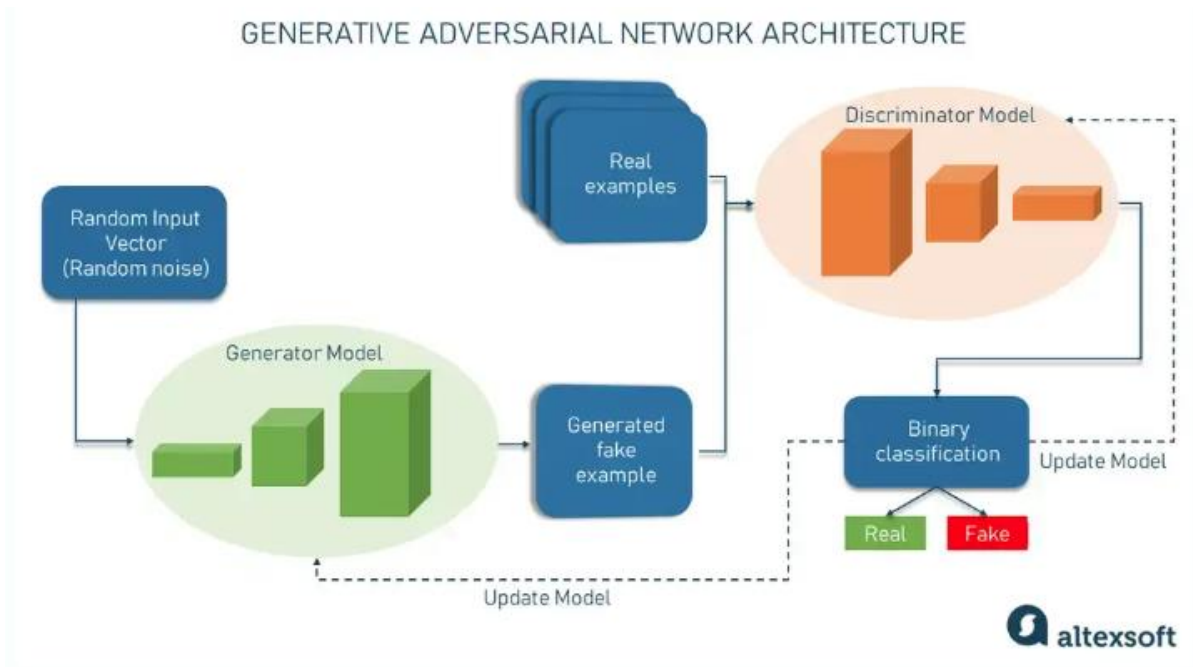
(AltexSoft, 2023)

Where do the generative image AI models get their data from?

As mentioned above, AI models must be trained using datasets in order to render results, and must be trained on exceptionally large amounts of data in order to produce more accurate results. Training data can be sourced both internally and externally from third party sources. An example of internal sourcing is Spotify's AI DJ ,as it tracks your listening history and in turn generates a playlists. Some examples of external or third party sourcing are from vendors who sell large amounts of data, such as Reddit, and open data sets provided by entities such as the government, research institutions, and companies for commercial purposes. Companies may also use internet scrapers to obtain data but this runs a higher risk as it can infringe on copyrights (*What data is used to train an AI, where does it come from, and who owns it?*, n.d..)

Who owns this data?

Data is not necessarily owned, instead different rights may be attached to different types of data and the owner of those right may choose to restrict the use of data by third parties. While various types of data is used in different types of generative AI models, we will focus solely on issues surrounding image rendering. Data use may have various blurred lines, but there some protections that come with it in terms of data being used without authorization resulting in infringement of copyright and database right. Copyright is most relevant as it is attached to most human-created content such as text, images, video, audio, or other means of artistic works and can constitute as infringed if all or a substantial part of the work in question is copied. Databases also have rights that protect their data from being extracted from said database without permission from the owner. An example of this is Getty Images as they have commenced legal proceedings in both the UK and US against Stability AI claiming that the use by Stability AI's "Stable Diffusion" generative AI using their images within their training data sets is copyright infringement (*What data is used to train an AI, where does it come from, and who owns it?*, n.d..)

Why do images look like that?

As most people who have attempted to render images in a generative tool, one can see there are still many drawbacks and areas for improvement within these tools. Most notoriously is the AI struggle with hands.
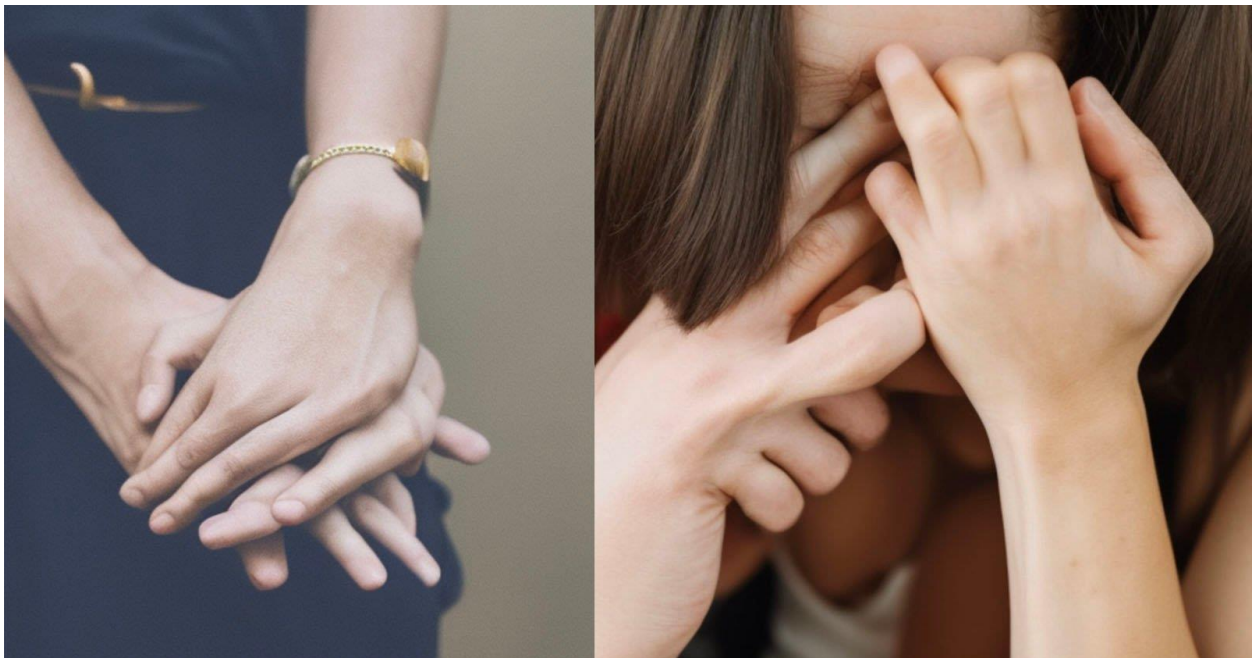
(Growcoot, 2023)

This tends to happen for various reasons from how people hold their hands in many different positions in source photos and their lack of presence of visibility in source photos used to train these models, to simply the technology behind it. Other examples that highlight the pitfalls of image generating involve texture and lighting resulting in blurry or jumbled images. Due to AI's lack of perspective it has a hard time with light placement and continuity of different textures, say from skin to hair and uses a medium to bridge the two, resulting in the flawed images.

(Edwards, 2023)