Generative Machine Learning: An Introduction to Models and Applications

Name: Egamov Giyos Gofur Ugli
Student Number: 2021040043
Department: Computer Engineering
University: Chungbuk national university

Introduction to Generative ML

Generative Machine Learning refers to a class of models that learn the underlying patterns or distribution of data so they can **generate new, original data** that is similar to the training examples. In contrast to discriminative models that focus on distinguishing between classes or predicting labels, generative models actually *create* data. In simple terms, **generative models can synthesize new samples (e.g., images, text) resembling the training data, whereas discriminative models classify or differentiate existing samples** . This fundamental difference means generative ML can do things like produce a realistic-looking photo of a person who doesn't exist, or write a paragraph of text in the style of Shakespeare, rather than just telling us which category a given input belongs to.

Why is generative ML important? It unlocks creative and practical applications that were previously impossible with traditional ML. By learning data distributions, generative models enable AI to **paint, write, compose music, design products, or even assist in scientific discoveries** – all by generating new content rather than analyzing something pre-existing. For example, a generative model can invent new artwork or help simulate possible chemical molecules. These capabilities have broad real-world use: imagine AI systems that generate lifelike images for graphic design, or chatbots that produce human-like dialogue. Such models power familiar tools like **chatGPT for text generation and deepfake or art generators for images**, showcasing how generative ML is transforming creative industries and everyday technology.

Core Principles and Learning Foundations

Learning Data Distributions (P(X) and P(X,Y))

Learning a data distribution means capturing how the data is spread out in terms of probabilities. For unlabeled data, models learn **P(X)**, the distribution of the input data alone. This is like understanding what a "typical" example looks like. If data has labels, a generative model can learn the **joint distribution P(X, Y)**, which captures how inputs and labels occur together . Learning P(X,Y) is like modeling the chances of each input and its label happening together in the data. With that knowledge, the model can generate new input–label pairs that follow the same patterns.

Likelihood Estimation

A key method to learn these distributions is **likelihood estimation**. Likelihood is a measure of how well the model explains the observed data. The model has parameters (like knobs) that it adjusts during training. It tweaks these knobs so that the actual training examples become as *likely* as possible under its model . This is called *maximum likelihood estimation*. For example, imagine a chef tweaking a recipe: each change to the ingredients makes the dish closer to the target flavor. The chef keeps adjusting until the dish exactly matches what they want. Similarly, the generative model adjusts itself until it would "cook up" the actual training data.

Latent Space

Many generative models use a **latent space**, which is like a hidden, compressed representation of the data . In latent space, complex inputs (such as images or sentences) are encoded as shorter, abstract codes. Each point in this space represents key features of an example, and similar inputs end up with similar codes. You can think of latent space like a map: it compresses each image or sentence into a small number of coordinates that capture its essence. For instance, a model might encode faces into a space where one direction controls “smile intensity” and another controls “head tilt.” This compact code makes it easier for the model to learn and generate new data.

Generating New Data (Sampling)

Once the model has learned the data distribution (and its latent space), it generates new examples by **sampling**. Sampling means picking random points from the learned distribution or latent space and converting them back into real data (for example, images or text). For example, a model trained on dog images can sample a point in latent space and decode it to create a new, realistic-looking dog image. In fact, one can say: *“If we sample* x_{\text{new}} *from* p(x)*,* x_{\text{new}} *should look like a dog.”* . A helpful analogy is to imagine a dog-shaped cookie cutter: each time you press it into dough you get a new dog-shaped cookie. All the cookies share the same basic dog shape but each one is slightly different. Similarly, sampling from the model yields many different outputs that all follow the learned pattern.

Types of Generative Models

Generative models come in various architectures and flavors. Three prominent types are Generative Adversarial Networks, Variational Autoencoders, and Generative Pre-trained Transformers – each with different mechanisms and use cases:

- **Generative Adversarial Networks (GANs):** A GAN pits two neural networks against each other in a creative duel . One network, the *generator*, tries to create fake data (e.g. an image) that looks real; the other, the *discriminator*, tries to tell apart the generator’s fakes from real data. They train together in an adversarial game: the generator improves at fooling the discriminator, while the discriminator sharpens its ability to catch fakes. Over time the generator learns to produce very realistic outputs, to the point that the discriminator can no longer tell fake from real . GANs are especially famous for image generation – for instance, producing photorealistic human faces. A striking example is the website *“This Person Does Not Exist,”* where each refresh shows a new human face completely dreamed up by a GAN (NVIDIA’s StyleGAN) . GANs have been used for creating artworks, realistic avatars, upscaling images, and even deepfakes. Their ability to hallucinate vivid, high-quality data makes them powerful, though training them can be unstable (the two networks must carefully balance each other).
- **Variational Autoencoders (VAEs):** A VAE is built on an **encoder-decoder** framework. The encoder network compresses input data (say an image) into a latent code, and the decoder network reconstructs the data from that code. Unlike a standard autoencoder, a *variational* autoencoder doesn’t produce a single fixed code for an input – instead it maps inputs to a **probability distribution in latent space**. In other words, VAEs encode each input as a range of possible latent values (a mean and variance), not one point . During generation, we can sample a random latent vector from this learned distribution and feed it to the decoder to get a new output. This stochastic approach forces the model to learn a smooth, continuous latent space that can be sampled for novel outputs. The result: VAEs can **synthesize new data that resembles the original dataset** (for example, generating new faces that look similar to those in the training set) . They tend to produce slightly blurrier images than GANs, but are more stable to train and have an explicit probabilistic foundation. VAEs are useful not only for image generation but

also for tasks like anomaly detection (since they learn a normal data distribution, anomalies stand out as poorly reconstructed) and data compression.
- **Generative Pre-trained Transformers (GPTs):** GPT is a family of generative models for **text** (and other sequence data) based on the Transformer architecture. Developed by OpenAI, GPT models are large language models that have been *pre-trained* on enormous text corpora and can generate human-like text . The core idea is to use the transformer's self-attention mechanism to model language: GPT is trained to predict the next word in a sentence given all previous words (this is called an autoregressive language model). Through this process on billions of sentences, it learns grammar, facts, and even some reasoning. **When you prompt a GPT model with a piece of text, it continues the text by predicting one word at a time, each time choosing the most likely next word based on the patterns it learned .** Despite the simplicity of next-word prediction, this yields impressively coherent and contextually relevant sentences. GPTs (like GPT-3 and GPT-4) can thus generate essays, answer questions, write code, or have conversations *ChatGPT-style*. They are called "pre-trained" because after the general training on internet text, they can be fine-tuned for specific tasks. These transformer-based generative models have revolutionized NLP by producing text that often reads as if a human wrote it, and they serve as the backbone for many modern chatbots and language-based AI services .

Architecture basics and differences: All three models above are generative but use very different architectures. GANs use two *convolutional* networks playing a zero-sum game (great for high-resolution images), VAEs use encoder–decoder networks with a latent probability distribution (giving a principled handle on sampling and anomaly detection), and GPT uses the transformer architecture with self-attention (excellent for sequential data like language). In practice, **GANs often produce the most visually sharp images (good for realism) but lack an explicit probability model,** while **VAEs have an explicit latent variable model and training objective (maximum likelihood via reconstruction and KL-divergence) which makes them easier to optimize, though their images can be fuzzier** . GPT and transformer-based models, on the other hand, are specialized for sequence generation; they handle long-range dependencies in text better than RNN-based models, thanks to the self-attention mechanism. Each approach has its strengths: for instance, **GANs shine in image fidelity**, VAEs in learned representations and continuity of latent space, and **transformers in capturing language context**. Often, the choice of model depends on the data type and the desired trade-off between sample quality and training complexity.

Neural Networks Behind Generative Models

Generative ML leverages various neural network architectures as building blocks, each suited to different data types:

- **Convolutional Neural Networks (CNNs):** CNNs are the workhorses for image-based generative models. Their ability to recognize and compose spatial features (edges, shapes, textures) makes them ideal for vision. In generative models like GANs and VAEs dealing with images, CNN layers are used to **encode images into latent representations and to decode or generate images by upsampling latent features into pixel space**. For example, the generator in a DCGAN (Deep Convolutional GAN) uses *transposed convolution* layers to progressively enlarge a low-dimensional noise vector into a full image. CNNs ensure that the local spatial coherence (like adjacent pixels having related colors) is learned and preserved in generated images. The result is that generative CNNs can produce outputs with realistic local details – think of generated faces having plausible eyes, noses, and mouths in the right arrangement, thanks to convolutional feature maps.
- **Recurrent Neural Networks (RNNs):** RNNs (and their improved variants LSTMs and GRUs) specialize in sequential data, making them traditional choices for generative tasks like text or music generation. An RNN processes one element of a sequence at a time while maintaining a hidden state that carries context. In a generative setting, an RNN can be trained on sequences (for example, sentences or melodies) and then used to **generate new sequences by predicting one element at a time**. Classic

examples include character-level text generators: an RNN, after being trained on Shakespeare's works, can generate new "Shakespeare-like" text one character after another. It keeps track of context (past characters or words) in its hidden state, enabling it to produce output that follows a learned style or structure. While RNNs have largely been overtaken by transformers for complex language generation (due to issues like short-term memory and difficulty in parallelization), they laid the groundwork for sequence generation and are still useful for simpler or more structured sequence tasks. In music generation, for instance, an RNN can learn musical patterns and compose new tunes by outputting one note at a time in sequence.
- **Transformers:** Transformers are the state-of-the-art architecture for many generative modeling tasks, especially in NLP. A transformer model uses self-attention mechanisms to handle input data, allowing it to consider relationships between all parts of a sequence in parallel. This is extremely powerful for language, where a word's meaning can depend on words from far earlier in the sentence. **For generative modeling, transformers (like those in GPT) excel at language generation by looking at all previously generated words and deciding the next word based on learned attention weights**. They can capture long-range dependencies better than RNNs, which is why models like GPT-4 can generate paragraphs of coherent text or code. Transformers are not limited to text – they have been applied to image generation (e.g. Vision Transformers and transformer-based diffusion models) and other domains as well. However, their largest impact has been on language: by processing text in parallel and learning contextual relationships, transformer-based models produce remarkably fluent and context-aware outputs. In summary, the transformer's role in generative models is to enable **scalable and context-rich generation** – making it possible to generate long sequences (like an essay or a program) where global coherence matters, something difficult for older architectures to achieve.

Each of these neural network types plays a role in modern generative AI. Often they are combined: for instance, some image generation approaches use CNN decoders with transformer-like attention for global coherence, and video generation might stack CNNs (for frames) with RNNs or transformers (for temporal continuity). Understanding CNNs, RNNs, and transformers provides a window into how generative models learn to create – from the pixel-level details up to global structure and context.

Applications of Generative ML

Generative models have blossomed into a wide array of applications across industries. Here are some prominent application areas:

- **Text Generation and Chatbots:** One of the most visible applications is in generating human-like text. Large language models (GPT-based and others) can produce coherent paragraphs, hold conversations, answer questions, and even write code. For example, **OpenAI's ChatGPT is powered by a generative model that can engage in dialogue, explain concepts, and create content on demand**. Similarly, GPT-based systems are used for drafting emails, writing articles, translating languages, and summarizing documents. The ability to generate natural language has transformed virtual assistants and customer service via chatbots, making interactions more intuitive. Beyond chatbots, text generation is used in creative writing (AI co-authors), game narrative development, and accessibility (converting structured data into readable narratives).
- **Image Synthesis and Art Generation:** Generative ML can create images from scratch or from prompts, leading to a boom in AI-generated art and design. Models like **OpenAI's DALL·E 2 and Midjourney can turn a text prompt into a vivid image**, often with stunning originality. You can ask for "a medieval castle in the style of Van Gogh" and get a unique image visualizing that description. Artists and designers use these tools to brainstorm ideas or create graphics. Additionally, GAN-based models can generate incredibly realistic faces (as mentioned earlier) and other imagery – these power everything from entertainment (e.g. generating characters or scenes) to practical uses like creating synthetic datasets for training other AI (by generating varied images to augment real data). There are

also photo editing and enhancement uses, where generative models fill in missing details (image inpainting) or increase resolution (super-resolution). The creative industries have embraced generative art, with AI-generated paintings and illustrations featured in exhibitions and graphic design workflows.

- **Music and Audio Generation:** Generative models are composing music, designing sounds, and even synthesizing realistic speech. In music, AI can learn from a library of songs and then **generate new musical compositions** in a similar style. For instance, researchers and musicians have used RNNs and transformers to generate classical music or pop melodies, and projects like OpenAI's MuseNet and Jukebox can produce multi-instrument music or mimic the style of famous artists. There are already albums composed with the help of AI. On the audio side, models can generate human speech with natural intonation (text-to-speech systems using generative adversarial approaches for voice) or create sound effects. Generative AI has been used to make background music for video games or to allow virtual personas to sing. Even visual artists like Holly Herndon have trained AI ("Spawn") on their own voice to create new vocals . These applications demonstrate that generative ML isn't limited to vision and text – it's making an impact in auditory and creative domains as well.
- **Scientific Discovery and Simulation (e.g. Drug Design):** Generative models are accelerating innovation in science by simulating data and proposing novel designs. In **drug discovery**, generative neural networks can imagine new molecular structures that satisfy certain medicinal properties – essentially suggesting new candidate drugs that chemists might not have thought of. For example, variational autoencoders and graph-based generative models have been used to generate novel chemical compounds with desired characteristics, like binding to a target protein. *Deep generative models have increasingly been used in chemistry for* ***de novo*** *design of molecules with desired properties* , offering a way to explore the vast space of possible compounds more efficiently. In physics or engineering, generative models can simulate possible experimental data or help design new materials by generating hypothetical material structures. Another famous example is **DeepMind's AlphaFold**, which, while not a generative model in the traditional sense of sampling random data, uses deep learning to *predict* protein structures – effectively "generating" accurate 3D configurations of proteins from amino-acid sequences, a task crucial for biology and drug development. (AlphaFold's AI-predicted protein structures have been a huge leap for science .) More broadly, generative models assist in any scenario where "imagination" of new possibilities is useful: creating synthetic training data for autonomous vehicles (simulating rare driving scenarios), generating virtual environments for robotics, or proposing new designs in architecture and engineering. By modeling complex distributions, generative ML serves as a kind of **virtual R&D assistant**, churning out ideas that humans can then test and refine.

Use Cases from Industry and Research Labs

Many leading tech companies and research labs have invested in generative AI, yielding some groundbreaking projects and products:

- **OpenAI:** OpenAI has been at the forefront of generative ML. Notably, they developed the **GPT series** (Generative Pre-trained Transformers) for language, which underpins ChatGPT and various other AI writing tools. These models have demonstrated how far AI-generated text has come, handling tasks from coding help to creative writing. OpenAI also created **DALL·E**, a generative model that produces images from text descriptions, showcasing multimodal creativity. These tools highlight OpenAI's mission to build general-purpose generative agents that can assist with a wide range of tasks – writing emails, designing visuals, tutoring, etc. The success of GPT models in particular has spurred an industry race in larger and better language models.
- **NVIDIA:** NVIDIA's research has produced state-of-the-art generative models in the visual domain. A prominent example is **StyleGAN**, a GAN architecture for incredibly realistic image generation (famous for those photorealistic human faces). StyleGAN introduced ways to control the "style" of generated images at different levels (like overall face shape vs. fine details), enabling high-quality and customizable outputs. NVIDIA has used generative models to create **synthetic data** for training AI – for

instance, generating thousands of varied images of objects or scenes to improve computer vision models. They also integrate generative techniques into graphics and animation, like using AI to generate realistic facial animations or to imagine how a scene would look in a different weather or time (useful in simulating data for self-driving car AI). NVIDIA's focus shows how generative ML can enhance both AI development and graphics technology, often leveraging their powerful GPUs for the heavy computation such models require.

- **Google DeepMind:** DeepMind (now part of Google) has applied generative and advanced predictive models to scientific problems. The standout is **AlphaFold**, which isn't a generative model that produces new training-like data, but rather a model that *learns the patterns of protein folding*. It solved the decades-old challenge of predicting protein structures from sequences, essentially generating likely 3D configurations of proteins with remarkable accuracy . This breakthrough (predicting ~200 million protein structures) is revolutionizing biology and medicine by providing insights into proteins quickly. DeepMind has also explored generative models in other areas – for example, using generative models for **game playing** (to imagine next moves), and they've developed generative agents in models like **Dreamer** for planning. Additionally, Google Research (Brain team) has its own generative projects: **Imagen** (a text-to-image model similar to DALL·E), **MusicLM** (a generative model for creating music from text descriptions), and more. These endeavors illustrate the wide impact of generative ML, from fun applications to profound scientific tools.
- **Kakao Brain (Korea):** As a Korean example, Kakao Brain (the AI research arm of tech company Kakao) has been developing large-scale generative models to keep pace with global AI trends. For instance, Kakao Brain released **Karlo**, a text-to-image generation model similar to DALL·E that can create images from prompts in Korean or English . They also built **KoGPT**, a Korean-language GPT model, to enable high-quality Korean text generation . Kakao's efforts culminated in launching services and APIs for image generation and text completion tailored to the Korean language and cultural context. This reflects a broader trend of localized generative AI – ensuring that non-English languages and diverse cultural content are well-served by generative models. Kakao Brain's work has put Korea on the generative AI map, showing that innovation is happening worldwide. Their open-sourcing of some models and funding of startups (e.g., the Karlo 100X project) further spurs generative AI development in the region.

GAN-Based Services and Technical Innovations

Generative Adversarial Networks, in particular, have sparked a range of consumer-facing services and creative tools. Some notable GAN-powered (or generally generative) services include:

- **StyleGAN and "This Person Does Not Exist":** NVIDIA's *StyleGAN* algorithm, mentioned earlier, became famous for the website **ThisPersonDoesNotExist.com**. Every time you visit the site, it shows a hyper-realistic portrait of a person who isn't real – *entirely generated by a GAN*. The underlying technology uses StyleGAN to synthesize lifelike faces by sampling different points in the model's latent space . StyleGAN's ability to generate convincing human images has been leveraged in various apps and games. Beyond faces, the technique has been adapted to generate art, landscapes, and more. The innovation here is not just the realism, but also the control: StyleGAN allows tweaking of visual attributes (like age, hairstyle, lighting) without retraining, by navigating the latent space. This has made it a popular foundation for **GAN-based image editing tools** and artistic applications where users want to generate new images with specific characteristics.
- **Runway ML (Video and Image Editing):** Runway ML is a creative suite that brings generative AI to artists, designers, and video editors without requiring coding. It incorporates numerous generative models under the hood. For example, Runway ML offers **video editing capabilities powered by generative AI**, such as background removal, style transfer, and even generating short video clips from text prompts. It recently gained attention for its text-to-video model Gen-2. As a platform, Runway demonstrates how generative models can be packaged into user-friendly tools. One can type a

description and get a brief AI-generated video, or use a GAN to modify a segment of a video (like turning a day scene into night). In essence, *Runway is a multimodal generative AI toolkit that can create high-quality images and videos and apply AI effects in real time* . It's been used in film production and design to quickly prototype visuals or achieve effects that would be hard to do manually. The advent of such tools means individual creators can leverage state-of-the-art generative models for their projects with ease.

- **Reface App (Face Swapping):** Reface is a popular mobile app that lets users swap their face into photos, GIFs, or movie scenes, often with uncanny realism. It utilizes generative models (akin to GANs) and face tracking technology to pull off this face-swapping trick in a matter of seconds. Essentially, you provide a selfie, and Reface will **generate images or videos where your face replaces that of a celebrity in a famous movie clip**, for example. Under the hood, *Reface uses GANs and advanced face alignment algorithms to produce remarkably lifelike face swaps in real time* . What used to require Hollywood-level visual effects can now be done by an app, thanks to generative AI. Reface's success also highlights some challenges – while fun, the tech is similar to what's used in deepfakes, raising questions about misuse. Nonetheless, as a service, Reface showcases how GAN-based innovation can go viral, giving everyday users a taste of AI-generated magic (seeing yourself singing a pop song or acting in a movie scene).
- **LG's EXAONE (Korean GAN-based Service):** LG AI Research in Korea has developed **EXAONE**, a large-scale AI model and platform that is part of the country's push into generative AI. EXAONE is not a single-purpose GAN but rather a comprehensive foundation model (with large language model capabilities, and potentially multi-modal features). It's notable for being one of the first major *Korean-developed* generative AI models. LG has iterated on EXAONE (from version 1.0 to 3.5), aiming to deploy it across LG's businesses and beyond. For example, they created **ChatEXAONE**, an AI assistant for internal use, based on EXAONE 3.5. *EXAONE 3.0 was introduced in 2024 as a huge advancement in large language models, and LG made a landmark move by open-sourcing it – the first time a Korean company open-sourced a large-scale model of this kind* . This openness indicates a drive to foster an AI ecosystem. EXAONE is designed to handle Korean and English, and it's been trained on diverse domains (from patents to chemistry papers). While it's more of a foundation model than a simple GAN service, we include it here as a **Korean innovation in generative AI**. LG's platform can be adapted for various purposes: generating text (like an enterprise ChatGPT), analyzing documents, possibly creating images or designs (if multimodal data is involved), etc. The development of EXAONE underlines that generative AI is a global endeavor, with companies like LG investing in their own versions of the technology to serve local needs and compete internationally.

Conclusion

Generative Machine Learning has rapidly grown from a niche research topic to a centerpiece of modern AI, bringing along a mix of excitement and caution. In summary, **generative models learn the patterns of a dataset so well that they can create new examples – a feat with profound implications**. We've seen that these models are important not just for academic curiosity but for real-world impact: they enable AI to be creative and proactive, producing content and ideas. This generative capability is revolutionizing fields from art to science. For instance, AI can help authors write stories, assist artists in visualizing concepts, or propose new molecular structures for pharmaceuticals. Such benefits come with the advantage of automating content creation, augmenting human creativity, and exploring infinite what-if scenarios in simulations.

However, along with the **benefits**, there are clear **limitations and challenges**. Generative models often require very large training datasets and computational resources, especially the likes of GPT-4 or image diffusion models – this makes them accessible mainly to well-funded organizations. They can also be *imperfect*: a text model might produce plausible-sounding but incorrect statements, and an image model might generate distorted hands or artifacts. Ensuring quality and reliability is an ongoing challenge. Moreover, generative models typically lack true understanding – they pattern-match rather than reason, which means they can sometimes produce outputs that are nonsensical or reflect biases present in training data.

Finally, it's important to address the **ethical issues** that arise. The power to generate content indistinguishable from real (like deepfake videos or realistic fake news text) raises concerns about misinformation, fraud, and privacy. There are questions of intellectual property – if a model trained on thousands of artworks creates a new painting, is it original or a remix of those artists' styles? Bias in training data can lead to biased outputs, potentially exacerbating social biases if not checked. Developers and society at large are now grappling with how to use generative AI responsibly: proposals include watermarking AI-generated media, setting usage guidelines, and improving model transparency. In spite of these concerns, the consensus is that generative ML, when guided by thoughtful policy and ethics, will continue to be a positive force, unlocking creativity and solutions we've only begun to imagine.

Generative Machine Learning represents a significant shift in how we conceive of AI's role – from passively analyzing data to actively creating. As research and development continue, we can expect even more sophisticated generative models that produce higher quality outputs and integrate better controls. For an undergraduate or anyone entering this field, understanding generative ML is not just academically fascinating but also increasingly essential, as these models are poised to become foundational tools in technology and industry. The journey of generative AI is just beginning, and it carries the promise of AI systems that are collaborators in creativity and discovery, expanding the horizons of what machines (and humans together with machines) can do.

References: ibm medium genaitoday.ai techcrunch.com d2l.ai