

---

## CHAPTER

# 1

# AI in a Nutshell

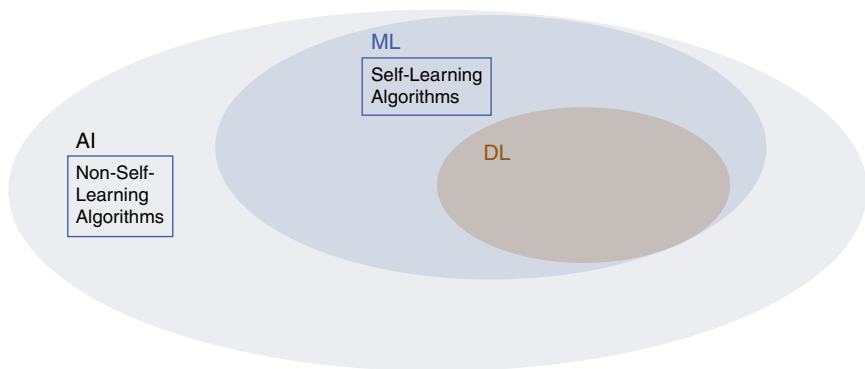
---

No other field of technology has such inconsistent jargon as artificial intelligence (AI). From mainstream media to tech influencers to research scientists, each layer of media has contributed to that confusion. In order of their degree of contribution and frequency, I observed mainstream media simplifying and misusing terms consistently, tech influencers misunderstanding the tech in-depth, and even some research scientists overcomplicating their model findings with fancy terms. By no means do I intend to criticize research scientists. They are the backbone of everything discussed in this book. Their work offers solutions to a plethora of problems, making AI the umbrella term for almost every intelligent problem. However, its interdisciplinary nature, the rapid advancements in this space, and AI's general complexity make it already difficult to gain a clear understanding of this field. I am convinced that consistent and clear language would help to understand this topic area.

We can see two broad classes in AI: generative AI, the subject of this book, and discriminative AI. The latter is the traditional and better-known part of AI. Before delving into both AI classes, let's take a moment to understand the broader picture of AI, machine learning (ML), deep learning (DL), and the process of training models, to avoid getting ahead of ourselves.

## What Is AI?

Even though AI includes a broad spectrum of intelligent code, the term is often incorrectly used. Figure 1.1 shows how AI, ML, and DL are related. ML, a part of AI, learns from data. DL, a deeper part of ML, uses layered setups to solve tougher problems. Non-self-learning programs like expert systems don't learn from data, unlike ML and DL. We'll explore these more next.



**FIGURE 1.1** The relationship between AI, ML, and DL

### How AI Trains Complex Tasks

AI can perform tasks ranging from predefined expert answers, also known as expert systems, to tasks that require human-level intelligence. Think about recognizing speech and images,

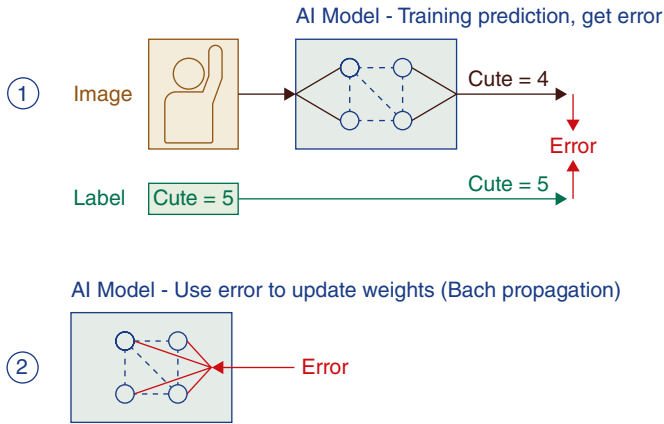
understanding natural language processing (NLP), making sophisticated decisions, and solving complex problems. For tasks like this, the AI has to train on a respective dataset until it is able to perform the desired activity as well as possible. This self-learning part of AI is referred to as *machine learning* (ML). Because most of the interesting applications are happening through machine learning in one way or another, and to keep it simple, we use AI and ML interchangeably.

To make it tangible, we are designing an AI system that rates the cuteness of cats from 5 (absolutely adorable) to 1 (repulsively inelegant). The ideal dataset would consist of pictures of cute kittens, normal cats, and those half-naked grumpy cats from the Internet. Further, for classifying pictures in a case like this, we would need labeled data, meaning a realistic rating of the cats. The model comes to life through three essential steps: training, validation, and evaluation.

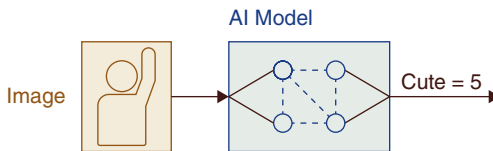
In training, the model looks at each picture, rates it, compares it with the actually labeled cuteness of the cat, and adjusts the model's trainable parameters for a more accurate rating next time—much like a human learns by strengthening the connections between neurons in the brain. Figure 1.2 and Figure 1.3 illustrate training and prediction, respectively.

Throughout the training process, the model needs to make sure training goes in the right direction—the validation step. In validation, the model checks the progress of the training against separate validation data. As an analogy, when we acquire a skill like solving mathematical problems, it makes sense to test it in dedicated math exams.

After training has been successfully completed and respective accuracy goals have been reached, the model enters the prediction or evaluation mode. The trainable parameters are not being adjusted anymore, and the model is ready to rate all the cats in the world.

**AI Model in Training (2 Steps)**

**FIGURE 1.2** In supervised training of a ML model, two main steps are involved: predict the training data point, then update the trainable parameters meaningfully based on the prediction's accuracy.

**AI Model in Prediction**

**FIGURE 1.3** Prediction mode in a supervised ML model.

It is typical for a model in production mode that the accuracy gets worse over time. The reason for this could be that the real-world data changed. Maybe we are only looking at kittens and they are all cute compared to our training data. Retraining the model, whenever accuracy decreases or by scheduling retraining periodically, tackles the problem of a discrepancy between the data distribution of training data and evaluation data.

Perhaps you have a sense already that training AI models requires much more computing power than they need in prediction mode. To adjust its trainable parameters, often referred to as *weights*, we need to calculate the grade of adjustment carefully. This happens through a famous model function called

*backpropagation*. It entails the backward propagation of prediction errors—the learning from making mistakes in the training process. The errors are turned back to respective weights for improvement. This means that we go forward to predict a data point and backward to adjust the weights. In prediction mode, however, we don't adjust the weights anymore, but just go forward and predict. The function that has been trained through the training data is being applied, which is comparatively cheap.

## Unsupervised Learning

When ML models reach a certain complexity by having many computing stages, called *layers*, we enter the realm of deep learning (DL). Most of the cutting-edge applications are at least partially drawing their algorithms from DL. Algorithms are step-by-step instructions for solving problems or performing tasks.

The preceding example of rating the cuteness of a cat was simplified drastically and didn't tell the whole story. A relevant addition to this is that as we train on labeled cat pictures, with the label being the cuteness of the cats, we call this *supervised machine learning*. With labels, we provide guidance or feedback to the learning process in a supervised fashion.

The counterpart for supervised ML is called *unsupervised machine learning*. The main difference between them is that in unsupervised ML the training data is not labeled. The algorithms ought to find patterns in the data by themselves.

For example, imagine you have a dataset of customer purchases at a grocery store, with information about the type of product, the price, and the time of day. In AI these attributes are called *features*. You could use an unsupervised clustering algorithm to group similar purchases together based on these features. This could help the store better understand customer buying habits and preferences. The algorithm might identify

that some customers tend to buy a lot of fresh produce and dairy products together, whereas others tend to purchase more processed foods and snacks. This information could be used to create targeted marketing campaigns or to optimize store layout and product placement.

Comparing the performance of unsupervised learning applications to that of supervised learning applications is akin to contrasting boats with cars—they represent distinct methodologies for addressing fundamentally diverse problems. Nevertheless, there are several reasons why we reached success years faster with supervised than with unsupervised learning methods.

In supervised learning, the model is given a training dataset that already includes correct answers through labels. Understandably, this helpful information supports model learning. It also accurately outlines the AI model's intended objective. The model knows precisely what it is trying to achieve. Evaluating the model's performance is simpler than it is in unsupervised machine learning, as accuracy and other metrics can be easily calculated. These metrics help in understanding how well the model is performing.

With this information, a variety of actions can be taken to enhance the model's learning process and ultimately improve its performance in achieving the desired outcomes.

Unsupervised models face the challenge of identifying data patterns autonomously, which is often due to the absence of apparent patterns or a multitude of ways to group available data.

## Generative AI a Decade Later

Generative AI predominantly employs unsupervised learning. Crafting complex images, sounds, or texts that resemble reasonable outputs, like an adorable cat, is a challenging task compared to evaluating existing options. This is primarily due to the absence of explicit labels or instructions.

Two main reasons explain why generative AI is taking off roughly a decade after discriminative AI. First, generative AI is mostly based on unsupervised learning, which is inherently more challenging. Second, generating intricate outputs in a coherent manner is much more complex than simply choosing between alternatives. As a result, generative AI's development has been slower, but its potential applications are now visible.

Between supervised and unsupervised learning, there are plenty of hybrid approaches. We could go arbitrarily deep into the knick-knacks of these ML approaches, but because we want to focus on generative AI, it is better to leave it at that. If you want to dive deeper into the technicalities, I recommend the book *Deep Learning* (Adaptive Computation and Machine Learning series), by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (MIT Press, 2016), which covers ML and DL in great detail, laying the theoretical generative AI foundation. It is regarded as the best book in the space, which isn't surprising, given the authors. I will come back to those gentlemen later.

The AI landscape is vast and ever-expanding. In this book, I strike a balance between simplifying concepts for clarity and providing sufficient detail to capture the essence of recent AI advancements. To understand what generative AI is and its value proposition, we first have to understand the traditional part of AI, called *discriminative AI*.

## What Is Discriminative AI?

Discriminative AI models made headlines long before large language models (LLMs) like ChatGPT by OpenAI and image generation models like stable diffusion by Stability AI entered the stage. Since the term “artificial intelligence” was coined by John McCarthy in 1955, discriminative models have yielded great results, especially in the past 15 years.

Discriminative AI focuses on algorithms that learn to tell apart different data classes. They recognize patterns and features unique to each class, aiming to link input features with labels for the output. This way, they can effectively classify instances into predefined groups, making it easier to distinguish one class from another. Discriminative AI has found numerous applications in various domains, including NLP, recommendations, and computer vision.

In the field of NLP, discriminative AI is used to classify text data into different categories, such as sentiment analysis or topic classification. In the domain of recommendations, discriminative AI is used to predict user preferences and make personalized product recommendations. In computer vision, discriminative AI is used to recognize objects and classify images based on their content. The applications of discriminative AI are vast and diverse, and its impact on various industries is immense.

Looking at existing applications, discriminative AI generally has five main tasks: classification, regression, clustering, dimensionality reduction, and reinforcement learning. They are not crucial to be able to follow the book's thread, but it helps to understand them conceptually because then the term “discriminative” and what it means in the context of AI becomes apparent. Put simply, in one way or another, this part of AI is deciding, selecting, distinguishing, or differentiating on data or a problem at hand.

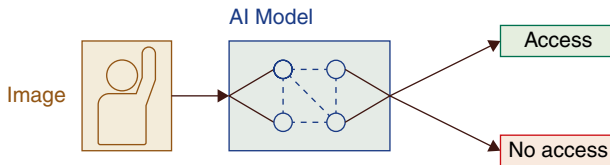
## Classification

The objective of classification is to accurately predict the class of new inputs based on prior training with labeled examples (Figure 1.4). This supervised learning process uses training examples accompanied by their respective class labels.

For instance, consider unlocking your phone with facial recognition. You initially show your face from various angles,



allowing the classifier model to learn your appearance. Advanced face recognition systems, like the iPhone's FaceID, quickly identify you due to their extensive pretraining and incorporation of biometric information to deterministically classify users. In essence, the model or system of models assesses your face and discriminates whether you belong to the “person with access rights” or “person without access rights” class.



**FIGURE 1.4** In ML, the concept of classification involves assigning data to one of a finite set of categories.

Classification has driven breakthroughs in diverse applications, including image classification, sentiment analysis, disease diagnosis, and spam filtering. These applications typically involve multiple processing steps and rely on deep learning techniques.

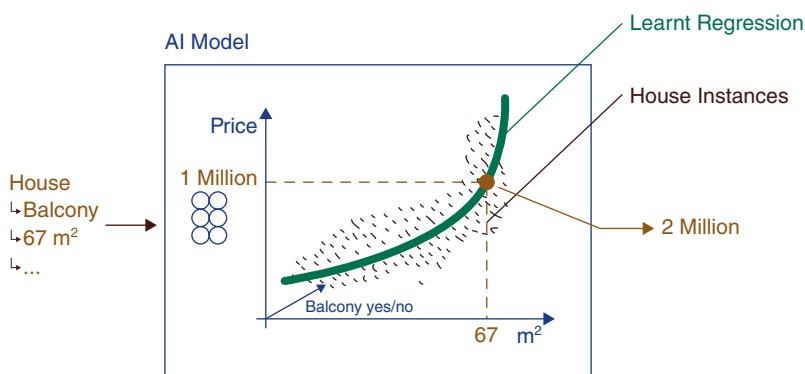
## Regression

A regression model in AI is designed to predict numerical values for new inputs based on data it has learned from a given problem. In this case, the output is not a class label but a continuous value. For example, imagine you want to buy a 100-square-meter apartment with a balcony in Munich, Germany. A real estate agent presents three similar apartments, priced at 2 million, 2.5 million, and 2.7 million euros.

You have three options: the naive approach, where you assume these three properties represent the market; the informed approach, where you estimate market prices by researching multiple offers; or the data science approach, which involves building a machine learning model to determine a fair price by

analyzing all available properties in the market with their price tags.

A well-trained regression model will give you a market-based and rational price, as it takes into account all the characteristics of apartments in the market (Figure 1.5), helping you make a more informed decision. By recommending a price, the model inherently has a discriminative nature.

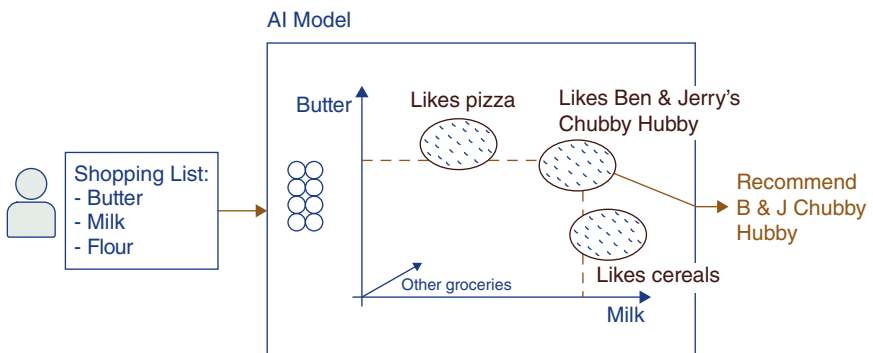


**FIGURE 1.5** In regression, data like house details go into the ML model, which then predicts its price based on these features.

## Clustering

As the name suggests, this application field in AI clusters data points. Be they people, groceries, or songs, based on a similarity measure, these items are grouped. By the way, you are being clustered all the time. For example, Internet ads are targeted to your digital persona, including your sex, age, IP address (which represents your location), and all other data ad-providing companies have collected about you. To cement it, if you use a web page that recommends songs like Spotify, movies like Netflix, and products like Amazon to you, then you have been clustered. In the success of big tech companies like those mentioned previously, clustering algorithms have played a crucial role, as they are the backbone of every recommendation engine.

In clustering tasks, the data comes without labels. For instance, there are no labels on our heads indicating “prefers Ben & Jerry’s Chubby Hubby.” Clustering models must identify patterns and groups autonomously, making it an unsupervised learning task. Moreover, the process of assigning items or personas to clusters is a decision-making aspect of discriminative AI. Figure 1.6 illustrates the conceptual operation of a clustering model. By analyzing other people’s behavior, it infers that individuals who purchase butter and milk might also prefer cereals. Adding soda to the mix increases the likelihood of a preference for Ben & Jerry’s Chubby Hubby.

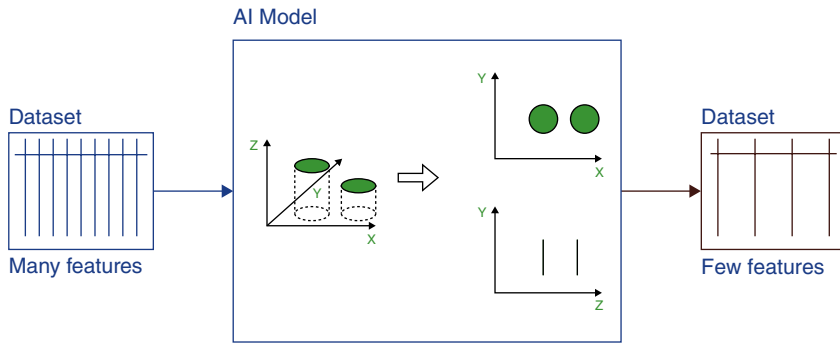


**FIGURE 1.6** Clustering model identifying buying patterns

## Dimensionality Reduction

Dimensionality reduction is not an application field of AI that is discussed much in mainstream media. It is rather research-heavy and often a means to achieve something greater, more efficiently.

Its primary purpose is to reduce low-information data, mainly making machine learning applications as effective as possible. By “low-information data,” I mean data that contains little to no meaningful insights to solve a problem. See Figure 1.7 for a visual representation.



**FIGURE 1.7** Dimensionality reduction

Imagine that you have an extensive recipe book with hundreds of recipes. Each recipe has several ingredients, and some of them are similar. For example, many recipes might call for salt, pepper, and olive oil. If you were to list all the ingredients used in the book, it would be a long list with many similar items.

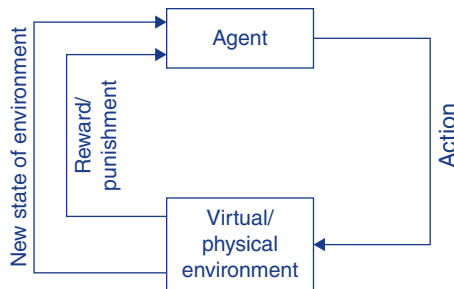
Now imagine that you want to make a simpler version of the recipe book that is easy to use on a daily basis. One way to do this is to group similar ingredients. For example, you could create a category called “seasonings” that includes salt, pepper, and other spices used in the recipes. You could also create a category called “cooking oils” that contains olive oil, vegetable oil, and so forth.

In the world of data science, the same thing happens. We might have a large dataset with many different features, and we want to simplify it to make it easier to work with. Dimensionality reduction techniques help us to do this by finding a way to represent the data with fewer features while still preserving essential information. They make it easier to analyze data, build models, or visualize data more understandably.

Naturally, the data is not labeled, and we don’t know up front which features carry relevant information. In an unsupervised manner, the models must learn to distinguish what low-information data can be modified or truncated and how. The models must decide or discriminate, indicating that we are in discriminative AI.

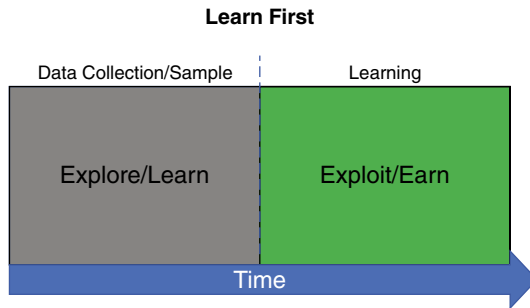
## Reinforcement Learning

Reinforcement learning (RL) models, typically called agents, learn from positive or negative consequences that their actions yield in real-world or virtual environments. A positive consequence is a reward, and a negative consequence is a punishment. In Figure 1.8, the agent executes an action in a virtual/physical environment, altering the environment (even if minimally), and receives a reward or penalty based on its stated goal. During the training phase of the RL model, initial emphasis is on exploration to identify available paths (e.g., for warehouse navigation), gradually shifting to an exploitation phase for efficient goal achievement (or technically, maximizing rewards), as indicated in Figure 1.9.



**FIGURE 1.8** Technical workings of reinforcement learning models

Virtual environments encompass a wide range of applications, from simulations for practicing real-world maneuvers to gaming experiences, and even stock market environments for trading agents. In gaming, AI has demonstrated remarkable super-human abilities, excelling in games such as *Super Mario*. When an RL agent acts in a real-world environment, it is probably a robot in a warehouse or Boston Dynamics's Atlas performing ninja moves. The agents acquire the ability to determine the optimal action in a given situation, positioning them as a component of discriminative AI.



**FIGURE 1.9** Exploration versus exploitation in RL training over time

Reinforcement learning has many exciting aspects, one of which is forming great synergies with generative AI. It was of little public interest for decades until its turning point in 2016, when AlphaGo by Google's DeepMind won a series of Go matches against the former world champion Lee Sedol. Go is a complex Chinese board game with a  $19 \times 19$  grid, and thus it has  $10^{172}$  possible moves. For comparison, there are  $10^{82}$  atoms in the universe. RL not only plays complex games exceptionally well but also delivers on a variety of tasks, ranging from autonomous vehicles to energy management in buildings. More on the powerful collaboration between RL and generative AI later.

Additionally, RL is helping to advance our understanding of the learning process itself, leading to new insights into how intelligence works and how it can be developed and applied.

## What Is Generative AI?

So far we have talked about discriminative AI, which can decide, distinguish, or discriminate between different options or continuous values.

Generative AI, however, is fundamentally different. It has the ability to generate all kinds of data and content. By learning the patterns and characteristics of given datasets, generative AI

models can create new data samples that are similar to the original data.

Recent advancements, such as the mind-blowing creations of Midjourney's image generation, the steps of video generation like Meta's Make-A-Video, and the conversational abilities of ChatGPT, have completely altered the way we view AI. It is a fascinating field that revolutionizes the way we create products and interact with data.

Generally speaking, generative AI models can perform three tasks, each with a unique and exciting set of applications.

### Data Generation

First, and it is the most obvious one, that they can generate all kinds of data, including images, videos, 3D objects, music, voice, other types of audio, and also text—like book summaries, poems, and movie scripts. By learning the patterns and characteristics of given data, generative AI models can create new data samples that are similar in style and content to the original.

### Data Transformation

The second task of generative AI is to perform data transformations. This means transforming existing data samples to create new variations of them. Transformations can reveal new insights and create appealing outputs for various applications. For example, you can transform winter pictures into summer pictures or day pictures into night pictures. Translating an image from one domain (for example, summer) into another (winter) is called a *domain transfer*. Image style transformation involves taking an image, such as a photograph of your garden, and maintaining the content (i.e., the garden) while altering its appearance to resemble the artistic style of, say, Monet's paintings. This process,

known as *style transfer*, is not limited to visual content like photos and videos but can also be applied to other data types like music, text, speech, and more. The essence of style transfer lies in preserving the original content while imbuing it with a distinct and recognizable, often artistic, flair.

Style transfer is more than just a delightful tool; it possesses the potential to significantly improve datasets for broader applications. For example, researchers from Korea and Switzerland have independently investigated the use of style transfer techniques to augment the segmentation of cancer cells in medical images using machine learning. This method, dubbed *contextual style transfer*, relies on the seamless integration of style-transferred instances within the overall image, ensuring a smooth and cohesive appearance—something that generative adversarial networks (GANs) are able to perform. In a fascinating study, Nvidia showcased a remarkable improvement in segmentation performance by incorporating synthetic data into the training set. This integration led to a leap from 64 percent to 82 percent in accuracy simply by augmenting the dataset, without modifying the machine learning pipeline in any way.

## Data Enrichment

As already indicated with style transfer, the third task of generative AI is to enrich datasets to improve machine learning models ultimately. This involves generating new data samples similar to the original dataset to increase its size and diversity. By doing so, generative AI can help to improve the accuracy and robustness of machine learning models.

Imagine we want to build a computer vision model that uses ML techniques to classify whether rare cancer cells are benign or malignant. As we are looking at a rare cancer type, it will be a small dataset to train on. In real-world scenarios, privacy issues are another data-diminishing factor. However, our neural net is



data-hungry and we can't get the most out of its power, landing at 64 percent classification accuracy. Through generative AI, rare cancer images can be generated to create a larger and more diverse training dataset for improved detection performance.

Overall, the capabilities of generative AI are truly remarkable, and the potential applications are vast and varied. AI limits are being pushed every day, not only by research but also by for-profit companies. This is especially true of generative AI.

If we zoom out further, we see that the overall concept of generative AI is even simpler. Models generate data based on some input. The complexity of the input can vary a lot. It could range from simple tasks, such as transforming a single digit like 6 into a handwritten image, to complex endeavors like applying domain transformations to a video.

### Under the Radar No More: Picking Up Speed

What we often observe, especially in AI, is that a new tech approach has early roots, but has been in stealth mode for a couple of decades. Once sufficient advancements transpire in a related tech domain, the dormant technology awakens, delivering substantial value in real-world applications. This is recognized as technological convergence.

**Deep Learning Tech Convergence with GPUs** The advent of deep learning, the underlying technology propelling fields such as computer vision and robotics, traces its roots back to 1967, when the first neural network, the multilayer perceptron, was conceived and introduced by two prominent Soviet scientists, Ivakhnenko and Lapa.<sup>1</sup> For numerous decades deep learning struggled to yield tangible business value and real-world

---

<sup>1</sup>A. G. Ivakhnenko and Valentin Grigor'evich Lapa, *Cybernetics and Forecasting Techniques*, American Elsevier Publishing Company, 1967.

applications. However, a transformative moment arrived with the emergence of graphics processing units (GPUs) at the onset of the 21st century.

GPUs first became popular in the gaming industry. In the late 1990s and early 2000s, video games became increasingly complex and required more processing power to render high-quality graphics and animations.

In the 1990s, GPUs were initially developed with the primary aim of providing specialized processing for intricate 3D graphics and rendering in video games and other computer applications. Firms such as 3DFX, ATI, and Nvidia spearheaded these advancements. The early 2000s witnessed another significant development for GPUs: the introduction of parallel processing, enabling multiple calculations to be executed simultaneously.

This ability to compute large amounts of data breathed new life into deep learning, allowing it to gain traction and experience a surge in research popularity. Leveraging GPUs' enhanced capabilities, researchers and practitioners accelerated deep learning's potential, sparking a multitude of practical applications. Today, it's unimaginable to train a robust machine learning or deep learning model without the assistance of GPUs.

Deep learning has reaped the benefits of other advancements as well. The Internet's growth and technological innovations provided abundant data for training models, while committed researchers and research, in general, led to numerous breakthroughs in deep neural networks. This progress extends from convolutional neural networks achieving remarkable feats in image recognition to recurrent neural networks demonstrating advanced NLP capabilities. It's not just the researchers who are passionate about the subject; capital allocators and profit-driven companies have also invested heavily in the field.

Incidentally, it's worth mentioning that we are now seeing, and will likely keep seeing, a similar rise in interest in generative

AI. The growth of other areas, especially discriminative AI and computational power, along with the increasing amount of data, were crucial for generative models to evolve in the background.

Today, we see billions being invested in generative AI projects aimed at tackling a wide range of business and non-business applications, as long as people can imagine it. This growing focus on generative AI promises to bring even more transformative advancements in the near future, building on the foundation established by previous AI breakthroughs.

In today's attention economy, capturing the focus of individuals has become increasingly challenging, as attention itself is a scarce and valuable resource. The widespread adoption of the Internet, social media, and other digital technologies has led to an overwhelming influx of information and stimuli, all competing for our limited attention. Consequently, only groundbreaking technologies can truly stand out and capture the spotlight. For a long time, generative AI remained relatively obscure in this competitive landscape. However, recent advances and remarkable achievements have now propelled generative AI into prominence, showcasing its immense potential and securing its place at the forefront of technological innovation.

**Generative AI's Early Impact** Generative AI is still quite new, but its future effects are expected to be amazing, going beyond what we've seen so far. Its influence can be noticed in many areas, but it has mainly made a difference in three sectors: creative industries, gaming, and natural language processing.

**Creative Industries** Generative AI has made a lasting impact on creative fields like art. This technology enables artists to create unique and inventive digital artworks. By studying patterns and styles in existing art, music, and fashion, AI algorithms can produce new content that matches market trends and engages

audiences. In the world of music, these algorithms can generate original tracks or remix current ones, opening up fresh possibilities for both producers and artists.

The integration of generative AI has led to new business models in the creative industry, such as selling exclusive digital art or creating customized products using AI-generated designs. This growth has occurred alongside a technological convergence between AI and the rapidly expanding cryptocurrency landscape.

In the last eight years, the cryptocurrency world has seen incredible progress, with numerous coins quickly making some people wealthy and leaving others financially devastated. Decentralized finance and institutional adoption have drawn significant interest. However, the most far-reaching impact may come from non-fungible tokens (NFTs).

NFTs allow artists and creators to produce unique, verifiable digital assets, leading to a growing demand for imaginative, high-quality AI-generated art. While not the sole driving force behind advancements in image generation, the NFT market has undeniably accelerated progress in this area.

**Gaming Industry** The gaming industry has experienced a significant transformation due to generative AI, which has opened up possibilities for a variety of new game content, such as levels, characters, 3D objects, scenarios, and even entire quests. A notable example is Microsoft's Flight Simulator, which partnered with Blackshark.ai to generate a photorealistic, three-dimensional world from two-dimensional satellite images, covering the whole Earth.

The popularity of open-world concepts in gaming has encouraged many companies to adopt AI-generated content. Imagine AI algorithms that study player behavior and dynamically modify game difficulty or generate new content on the spot,

leading to personalized and engaging gaming experiences. Consider the potential of giving non-player characters (NPCs) AI-driven language models for more captivating and immersive interactions. These advancements could make returning to the real world a challenge.

By using generative AI to create in-game items and environments more efficiently, gaming companies can allocate more time and resources to concentrate on core aspects, ensuring the production of intriguing and original content. The future of gaming, fueled by generative AI, is set to be an exciting and immersive adventure for players.

**Natural Language Processing** The third impact vertical is not a single industry per se but rather many industries.

Generative AI can be used to generate new content such as text, summaries, or translations. Large language models are at the forefront of generative AI applications, with widespread impacts across various industries. LLMs can improve operational efficiencies by automating repetitive internal processes and accelerating innovation through customer feedback analysis, insights, and market research. These models can also improve customer experiences with concise answers and summaries available 24/7. The potential for managing knowledge is perhaps one of the most significant aspects of AI systems; organizations with specialized knowledge can offer their expertise in a tailored and concise manner to end users. Take the Mayo Clinic, for instance. Specializing in patient care, research, and education, the Mayo Clinic has amassed a wealth of data on medical conditions and treatments, such as patient records, research studies, and medical imaging data. They could create chatbots and virtual assistants that harness this data to provide expert guidance and advice to patients. By integrating these AI-driven tools into the Mayo

Clinic's website or mobile app, patients could access expert medical advice from anywhere around the globe.

Language models don't just generate language, but also code, music, poetry, stories, jokes, captions, summaries, translations, recommendations, and much more. The fields will further broaden, with LLMs providing innovative solutions for businesses and society.

Generative AI is immensely exciting as it will undoubtedly revolutionize how we create, consume, and process content across all aspects of our lives. As the technology develops, we can expect further paradigm shifts, leading to groundbreaking advancements in industries worldwide.