ALTAIR

A HUMAN'S GUIDE TO GENERATIVE AI

INTRODUCTION

The advent of generative artificial intelligence (AI) as a consumer product has sparked excitement in virtually every industry and brought a new level of awareness of machine learning's capabilities to the public. Now that generative AI is no longer confined to research labs and has entered the public square, it has sprouted into dozens of new applications and at least as many products. Generative AI could change how we interface with computers forever, and specific use cases are popping up constantly as Silicon Valley's collective imagination runs wild.

It's inevitable, however, that when a technology has this much hype, product designers, investors, and economic prognosticators alike get a little carried away in their excitement. While this is understandable, it's not all that helpful for businesses and individuals looking to determine what this tech means for them, in the real world.

In this eGuide, we'll provide a levelheaded, easy-to-understand overview of generative AI technology – what it is, where it came from, how it works, and what it can (and can't) do. We'll go over the considerations and responsibilities of any enterprise implementing generative AI, including how to access it, the risks and costs involved, and critically, the tasks at which generative AI excels. We'll compare generative AI's strongest use cases to problems that are still better suited to other kinds of machine learning algorithms and outline why it's important to pick the right tool for the job. Finally, we'll break down niche use cases industry by industry, highlighting opportunities in each field.

We are at the beginning of a wave of new technology. Generative AI can substantially accelerate the work of anyone who creates content. Similar to the invention and proliferation of the automobile, there are clear risks in generative AI, but I believe the benefits outweigh the risks enough that the question becomes how to use the technology safely – not whether to use it at all. At Altair, we take security, responsible methodologies, and privacy very seriously, as should everyone. We want to provide accelerators, but safe and secure ones.

Christian Buckner, Senior Vice President, Data Analytics and IoT, Altair

Table of Contents

Generative AI presents a dramatic leap forward in machine learning's capabilities, and tech leaders and expert consultants are making all sorts of claims about its future, forecasting business impacts in the billions and trillions of dollars and affecting hundreds of millions of jobs. Those are big numbers and serious claims. What's the real story?

The truth is that no one knows. But it's no exaggeration to say that generative AI has the capability to forever transform how humans interact with their virtual and digital environments and the way we think about knowledge production and dissemination. Generative AI also comes with real risks to consider alongside its potential benefits. Understanding how generative AI works will help contextualize what it can do, how to best use it for yourself or your enterprise, and how best to deploy it.

OVERVIEW OF ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, AND DEEP LEARNING

It's rare that a new technology comes along that gets everyone in the business world talking, but the wide release of generative AI has done just that. ChatGPT was announced in November 2022 and suddenly, it seemed that everyone was raving about the possibilities of generative AI. Advanced machine learning algorithms can now generate natural-sounding text and produce compelling imagery quickly and easily with minimal human involvement. Generative AI could profoundly shift the way ideas, data, and information are created and then shared, a domain that has required human intelligence since, well, the birth of language itself.

The conversations surrounding generative AI are reinvigorating longstanding debates about the power and purpose of human creativity, the role of technology in our personal and professional lives, and the very definitions of intelligence, learning, language, and art. Philosophical musings aside, the current enthusiasm for all things generative AI provides a great opportunity to dive into the complexities of a topic that will be indispensable to business and technology innovation for years to come.

First, let's clarify the computer science disciplines that have given rise to generative AI:
• Artificial intelligence (AI)
• Machine learning
• Deep learning

Each of these fields are interrelated but have unique characteristics, which is why it's important to distinguish them from one another. AI is the top-level category, machine learning is a sub-field of AI, and deep learning is a sub-field of machine learning. For a more visual example, think of this hierarchy as a set of nesting dolls in which AI is the largest doll that houses the others.

Data Science

Covers the practical application of advanced analytics, statistics, machine learning, and the necessary data preparation in a business context.
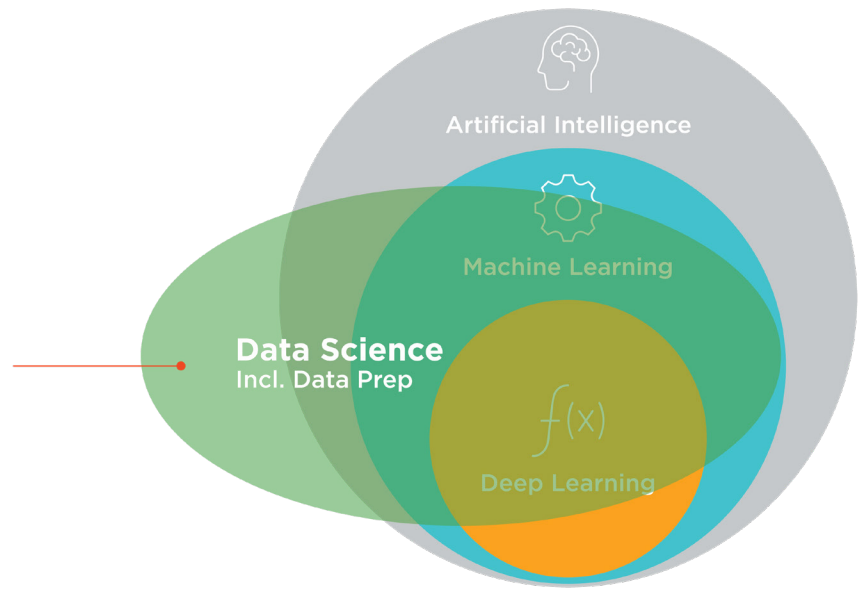
AI refers to any computer that can perform a task commonly associated with human cognition. Machine learning is a subfield of AI that focuses on training computers to find patterns in data and use what it learned to solve novel problems. Deep learning is a special group of machine learning methodologies whose internal structure, called a neural network, is meant to mimic the human brain. Neural nets are powerful problem solvers but often require special hardware due to their intensive computational demands.

Artificial Intelligence

We call any computer or computing system an "artificial intelligence" when it can perform tasks that have typically required human intelligence. That's a broad definition, but it's meant to encompass a wide variety of computing tasks such as:

- Speech and text recognition, interpretation, and generation,
- "Smart" assistants such as Apple's Siri or Amazon's Alexa,
- Search and recommendation algorithms for everything from Googling the best tent for your upcoming camping trip or browsing Netflix for your next binge-watch,
- Route optimization for drivers, shipping and logistics companies, and self-driving cars,
- Automated stock trading and investing,
- And many, many more.

All these AIs have two things in common: They work in a limited context, and their intelligence, such as it is, is applied to a narrowly defined problem. Often called "restricted," "weak," or "specialized," AIs, these AIs generally try to do one thing well, like mastering the rules of a board game or finding the best route from point A to point B. Even a smart assistant like Siri only focuses on a single function: recognizing human speech and turning it into text. Every other function relies on preprogrammed responses or fetching information from the internet based on the text the program has detected.

We're still a long way from "strong" AI, or artificial general intelligence (AGI). This might be what springs to mind when you hear "AI," the sort of machine intelligence we see in sci-fi movies, like HAL or Lieutenant Data. One of the hallmarks of AGI would be the ability to solve novel problems that it hasn't been trained to work on. To do this, it would need to use a full set of cognitive abilities on par with an adult's ability to reason, sense, feel, and ideate. The final hallmark would be self-awareness: the AI's knowledge of itself as an intelligence distinct from other intelligences, a conscious mind that understands its place in the world. Despite how similarly they're named, generative AI is **not** AGI.

Machine Learning

Machine learning is a sub-field of AI that focuses on pattern recognition and forecasting. Once given a sufficiently large and complex data set to study, a machine learning algorithm can identify patterns and then make predictions based on what it found. Machine learning programs use computational techniques and statistical rules to gather information directly from data without relying on mathematical models or predetermined equations. Basically, the algorithm can make predictions on an endless number of queries (within its scope of operation) without being explicitly coded to produce any of them, creating a degree of autonomy in computer behavior that would be impracticable with ordinary software engineering methods.

What sets machine learning apart from other data science fields, however, is its capacity to – we might have buried the lede a bit here – learn. If you define "learning" as "getting better at performing a task over time as experience is gained," then the algorithms really do learn.

Machine learning algorithms do this in two ways, broadly speaking: supervised and unsupervised. In brief, the "supervision" in a supervised model is just labeling attributes of the data, which can be any number of descriptive terms about the object the algorithm is trying to understand. These could be anything from email open rates for different customer segments to seismograph readings near an active volcano. Labels for the data are based on historical outcomes, like whether the customer unsubscribed or the volcano erupted.

As it studies the data set and finds patterns among the inputs and outcomes, the algorithm can begin to infer cause-and-effect relationships and then make guesses accordingly. Since there's historical data against which to compare the algorithm's predictions, it's easy to see how well the model is working, which means that supervised machine learning algorithms produce reliable, repeatable decisions.

Unsupervised learning algorithms, on the other hand, use data that hasn't been classified or labeled. In this case, the machine learning algorithm doesn't just predict outcomes, it also finds structure and patterns within the data on its own – without historical data guiding it towards desired outputs. Unsupervised algorithms are particularly good at figuring out how different data points might be grouped together into categories, and since the models are left to their own devices, they may surprise you with totally unexpected results. The downside is that users may not be able to identify why or how the algorithm reached its conclusions. Without a historical reference point to compare the model's outputs to, unsupervised learning is also harder to assess for quality and accuracy.

> Once given a sufficiently large and complex data set to study, a machine learning algorithm will identify patterns and then make predictions based on what it found.

Deep Learning

Deep learning is a sub-field of machine learning that teaches computers to solve problems using an internal software architecture inspired by the human brain. This structure is called an "artificial neural network," and it's made up of artificial neurons that work together to find a solution to the problems users pose to it. These artificial neurons are just very, very small software modules, called nodes, inside the larger deep learning program. Each node is a bit like an airline hub or metro station, with many connections to other nodes providing a lot of different routes for the information to take. The information travels from node to node, each one doing a tiny amount of computing and passing the information along to the next node based on the outcome as the model tries to find the best pathway for that query.

These nodes are arranged in consecutive layers: input, output, and some number of hidden layers in between. Information is taken in at the input layer, processed or analyzed, then passed to the next layer. The information is processed bit by bit in the hidden layers, a little more with each node it passes through on a path whose general shape is determined by the type of neural net it's in (but isn't known exactly by an outside observer). Finally, the output layer receives the bit of information after it's been plinko-ed around the neural network and presents the result in a legible fashion.

Many of the current market's most advanced AI applications – such as speech recognition, image identification and classification, and generative AI programs – are powered by deep learning.
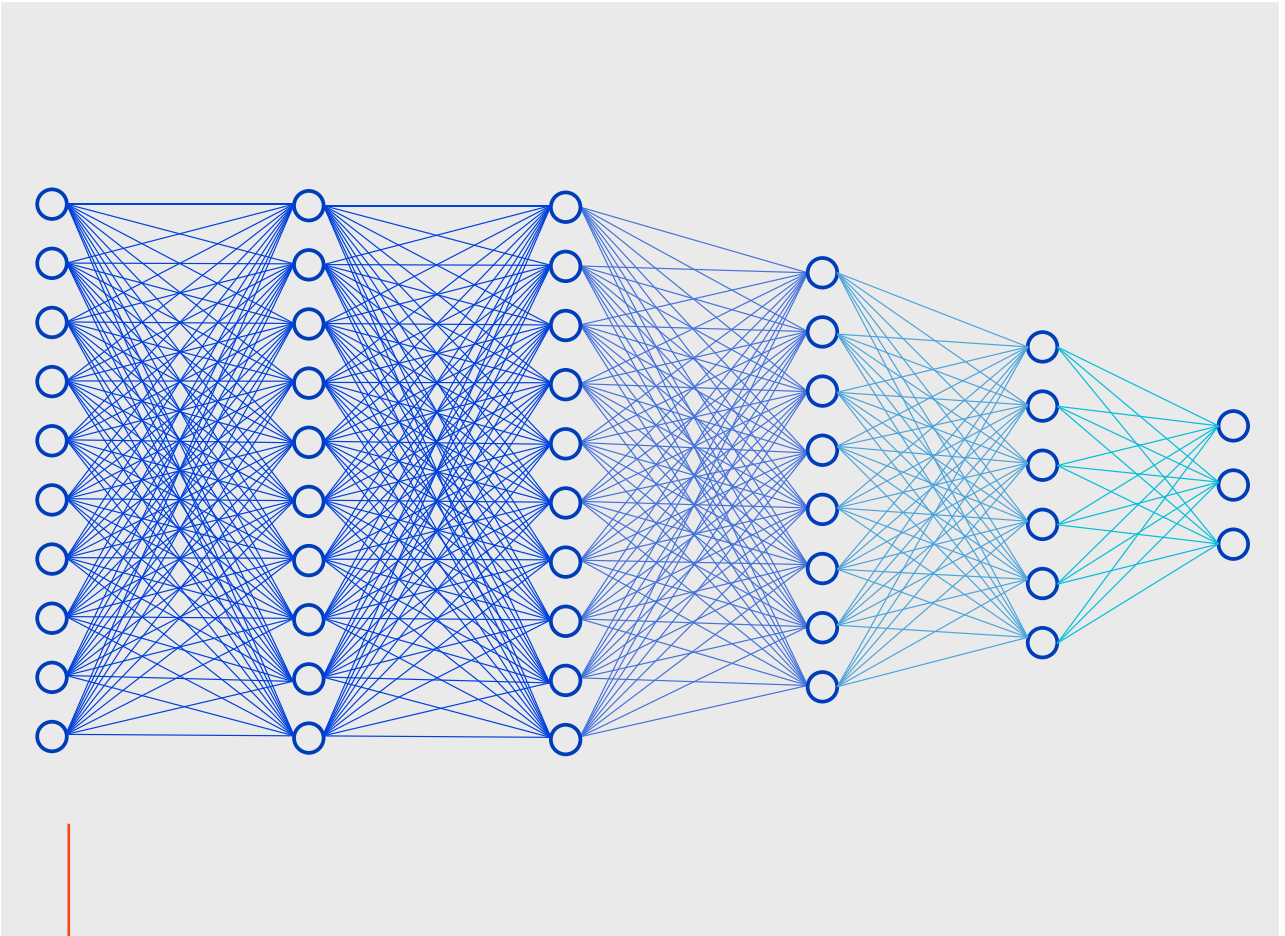
Deep learning models use an artificial neural network, a type of machine learning model inspired by the human brain, to solve harder problems more autonomously than any other technology can handle.

GENERATIVE AI

Generative AI is a class of machine learning algorithms that use neural networks to create text, images, and other content that is substantially different from anything it was trained on and much more complex than any previous machine learning model was capable of. These models are "generative" because their distinguishing feature is the creation of content with virtually infinite variety in form and substance. These models can mimic human visual and verbal outputs to an astonishing degree, making it possible for machines to produce text, audio, video, and imagery that can pass for human-created work.

Compare this with simpler machine learning models, what some specialists (including us at Altair) call "discriminative" machine learning, whose main purpose and strength is discovering patterns in its training data and using them to produce results to a query. Though discriminative models can do sophisticated data analysis and perform some pretty incredible feats, these models' range of outputs is much smaller than generative AI and their knowledge base is far more limited. For example, a discriminative model could predict whether Customer A or Customer B was more likely to churn. Generative AI could make that prediction as well, but also act on the churn forecast by writing personalized emails to Customer A and B convincing each person to stay, using messaging and incentives tailored to them and the probable cause of their churn. Both types of machine learning can solve for the probability of the customer leaving, but the outputs are of wholly different kinds.

Much like the internet, generative AI had been used in academic circles for quite some time before it became popular with the general public, and both innovations broke out of their niches for largely the same reason – a company figured out how to package and present it as a consumer product. Not to oversimplify a major technical achievement, but the release of ChatGPT in November 2022 brought two critical factors to generative AI applications: ease of use and ease of access.

Ease of access: Anyone with an internet connection can visit a website or use an app. There's no need for specialized equipment, software, or academic credentials to use Bard, Midjourney, DALL-E, and so on. Using generative AI tools machine learning suddenly became an everyday activity for the average netizen.

Ease of use: The current batch of generative AI tools are slick-looking and user friendly. ChatGPT is simplicity itself – just a text input on a webpage. The user types in what they want the program to do in plain language and – presto! – out comes the response. The user didn't have to write a line of code, let alone go through a decade of schooling in statistics, computer science, or software engineering.

Foundation Models

Foundation models are the base upon which specific generative AI applications are built. They're deep learning models trained on massive data sets – really huge, terabytes' worth of data – containing a wide array of unlabeled data. Most foundation models have been trained to predict the most likely word to fill in gaps in texts. Only after they have been fine-tuned can they be used to solve more specific tasks like parsing natural language, generating new text and images, and carrying on a conversation. Though there are a few different kinds of foundation models, the most important of these for most businesses will be large language models (LLMs).

Large Language Models

LLMs are trained for superior performance with text. LLMs can perform a variety of language-oriented tasks, first and foremost being content creation; they can also summarize text, parse natural language commands for tasks like searching a knowledge base or writing code, act as customer service chatbot agents, and so on. Because of the neural network that powers them, LLMs are capable of all the same classification, pattern recognition, prediction, and self-instruction functions as the (relatively) simpler machine learning models are. The big difference is their ability to create sophisticated, human-sounding content that is informed by, but not identical to, its training data. The level of complexity in their outputs is leaps and bounds higher than other machine learning tools.

LLMs work by studying a large body of text data scraped from all over the Internet. As the model trains itself, it creates new text passages and compares them to what it has in its data set. From there, the models use statistical analysis to judge how close these are to the reference data, learn what they can, and try again. This training process can take months, depending on how much data it's pulling from. The result is a model that can generate new texts most-likely word by most-likely word.

There are a few important names to know in the LLM arena. Google's jovially named BERT (Bidirectional Encoder Representations from Transformers) and OpenAI's GPT (Generative Pre-trained Transformer) were both released in 2018. Both use Transformer architecture, which just describes the particular arrangement of the nodes in the neural network. Transformer models were a leap forward from the generative adversarial networks (GANs) that had been dominating AI research for the previous decade and have proven to be especially adept with text. Google has since released Language Model for Dialogue Applications (LaMDA), which is focused on conversation, aiming to make chatting with an AI feel natural and easy. Microsoft and Nvidia collaborated on an open-source model called Megatron-Turing Natural Language Generation (MT-NLG).

You might have noticed there were only big players up there. That's because training an LLM – or any foundational model, really – is difficult, expensive, compute intensive, and time consuming. To produce convincing human-sounding text, AI engineers have found that ever-larger data sets are required to improve performance. OpenAI is rumored to have trained their latest model, GPT-4, with 170 trillion (yes, with a T) parameters. Even if that number is off, it's certainly larger than the previous version, GPT-3, which they have stated publicly was trained with 175 billion datapoints. According to their published research, the computational power required for building OpenAI's foundation models has doubled every 3.4 months since 2012. In order to keep up, the need for high-performance graphics processing units (GPUs) to handle the intense computing workloads has also skyrocketed – as has their price.

How much it costs to make, train, and run a foundation model depends on too many factors to give a general estimate, but it's a lot. Millions of dollars and months spent building and training. From the person-hours of the data scientists and other personnel training and reinforcing the model, to buying the top-of-the-line hardware and cloud instances necessary for the neural network to run, to paying the energy and water bills to power and climate-control the infrastructure, it's safe to say organizations need deep pockets to bankroll deep learning ambitions.

Even though building a new foundation model is out of reach for all but the most well-resourced enterprises, there are still plenty of ways to leverage generative AI tech in your business.

IMPLEMENTING MACHINE LEARNING AND GENERATIVE AI

There's a lot of excitement about what generative AI can do, but the truth is that discriminative machine learning models – which are cheaper and quicker to create – can handle many of the proposed use cases for generative AI (and LLMs in particular). No need to use a chainsaw to cut a birthday cake, right?

A few examples where a discriminative model will do:
- Anything to do with knowledge management, such as retrieving and synthesizing in-house data from unconnected and/or incompatible sources. Optimizing, rebuilding, and integrating legacy systems doesn't need an AI model. (In fact, this can be done without machine learning at all with Altair® Monarch®).
- Analyzing operations and logistics, like planning for inventory supply or staffing coverage.
- Optimizing pricing strategies.
- Recommendation engines for products, media, and more.
- Predicting customer behavior, including churn warnings and sales forecasting.
- Customer segmentation and persona identification.
- Sentiment analysis for call centers, product reviews, discussion forums, and similar venues.
- Simple image recognition and classification.
- Speech recognition for customer service, language learning, human-computer interfacing, and so on.

The breakthrough of generative AI is the diversity and substance of its outputs. The open-endedness, the malleability, and the novelty of what generative AI can produce is its unique and exciting contribution to the business world. The following are tasks where generative AI can really shine:
- Text generation and summarization (also known as augmented authoring and reading).
- Generative design and engineering, including the use of natural language commands to create models of physical objects, speeding up product development iterations, virtual prototyping, and testing.
- Synthetic data creation (using a computer simulation or algorithm to generate a wholly invented data set based on, but distinct from, a representative sample, which have the benefits of being anonymous, pre-labeled, and more distant from bias in the real-world data).
- Computer programming, writing, and debugging code.
- Chatbots with a good command of natural language and with customizable personalities for customer service.

The list can – and should – go on from here. This list just scratches the surface of what's possible, and there's still so much potential to be capitalized on. We at Altair want to encourage the data science field and the business world at large to dream bigger. Let's be bold in envisioning what the future of this technology could look like and how we use it.

CONSIDERATIONS AND RESPONSIBILITIES WITH GENERATIVE AI

There's a reason why we're emphasizing using the right machine learning tool for the job: There are costs and risks in using generative AI, and managing both is good business. Leaders and advocates should be informed so they can ensure proper risk-mitigation strategies are put in place. Read up on some of the snags users and enterprises could run into to better prepare to avoid them in the first place.

- Copyright infringement and copyrightability of outputs
- Reliability of content created
- Data privacy
- Model explainability and transparency
- Environmental impact
- Inherent bias and hate speech
- Cyberattacks and fraud

Intellectual Property and Copyright

We explained above how almost all LLM models are trained on a huge data set scraped from the internet. While a portion of that data would be in the public domain and could be considered "fair use," that's not necessarily the case for the whole data set. Foundation models trained on online content is almost certainly a violation of many, many copyrights, using and profiting from protected materials whose owners gave no consent nor received any compensation for their use. This is currently being litigated.

Copyright is an issue for outputs as well as inputs. For businesses hoping to build products or other proprietary offerings with generative AI tools, there will be issues protecting their new intellectual property. The U.S. Copyright Office has recently ruled that nothing that AIs produce substantially on their own is copyrightable. A copyright can only be granted to "a human author" with "ultimate creative control" over their work product. The copyright, trademark, and plagiarism issues (as both inputs and outputs of generative AIs) should be top of mind for any firm whose business is their intellectual property as they seek to monetize AI-produced work.

Reliability and Accuracy

There's no way around it: LLMs are designed to produce writing that looks and sounds like a human wrote it, not necessarily writing that's factually correct and supported by evidence. Often, generative AI models sound very confident. But sounding right doesn't mean being right. Generative AI models aren't sentient and they don't understand what they're saying the way humans do; these models just know that these sequences of words are very likely to appear in this order when those keywords are included. They can't distinguish truth from falsehood in the content they're creating, a significant pitfall that users need to be aware of. Generative AI chatbots are often incorrect, and worse than that, outright fabricate information and sources to support their claims. The consequences for taking an LLM at its word can range from mild embarrassment if the claim is fact-checked to making critical business decisions based on totally untrue information from an LLM prompt.

Privacy

If a user chooses to access the public-facing interface of a generative AI tool directly, they need to be aware that whatever they type into the prompt box becomes part of the model's training data and is used to further refine the product. This has already led to a leak of sensitive data from users at Samsung, resulting in a company-wide ban of the technology. Amazon has followed suit, as have many major banks. Any firm that works with sensitive and/or privacy-protected information should be leery of public generative AI tools. To avoid this, we recommend using generative AI models that are specially controlled for your organization.

Explainability and Transparency

It's in the very nature of neural networks to be black boxes – as a highly intraconnected unsupervised machine learning model, no one really knows the exact method the model uses to reach its conclusion. And now that these AIs are in the market as commercial products, business interests are driving companies to be more secretive regarding training data and neural net technology. Models of this kind are neither auditable nor explainable and thus don't provide any mechanism for accountability. If an unprofitable or even harmful decision is made because the AI recommended it, what can be done? That's a real issue for certain industries like healthcare, banking and financial services, government and policymakers, and any enterprise whose products touch human health and livelihoods.

Environmental Impact

Neural networks require a ton of compute power to train and maintain. The advanced GPUs needed to run the models consume a lot of energy, as do the climate control systems needed to cool the server stacks. Energy use equals carbon emissions and that leads to intensified climate change. (Not to mention that the electricity bill needs to be paid, too). Minimizing environmental impact is everyone's responsibility, and it's of particular concern to organizations that have made climate pledges or goals.

Bias and Hate Speech

Another consequence of LLMs being trained on data pulled unfiltered from the internet is that such a sample reproduces societal biases against marginalized groups. In fact, experts have noticed that sexist, racist, and other intolerant and hateful material is overrepresented in the source data because of the anonymous nature of online discourse. The guardrails that model owners have put on their products aren't always effective at stopping harm, since they don't address the root cause. Sadly, efforts to train the models not to generate hate speech have also caused harm, as they depend on a technique called reinforcement learning from human feedback. Powered by low-wage workers in Kenya and elsewhere, real people have to individually review offensive material in the dataset and label it descriptively so the model knows to exclude it.

Cyberattacks and Fraud

This is more of a heads-up for CISOs and IT leaders rather than a risk of using the product itself. Hackers and other bad actors are already using generative AI to deploy ransomware attacks, phishing scams, and other cybersecurity threats more quickly and easily than ever before. Countermeasures will need to be further refined starting immediately.

HOW TO ACCESS AND LEVERAGE GENERATIVE AI

Now that we have a map of the benefits and pitfalls of generative AI, we can start thinking about how to take advantage of all it has to offer while staying savvy to when a discriminative machine learning model might do the trick. The next question is: How does my enterprise access generative AI applications in a cost-effective, secure, and right-sized way?

Firms have four options. The easiest is to access a consumer-oriented version of one of the generative AI models currently available, like ChatGPT on the web or joining Stable Diffusion's Discord server. Given the issues outlined in the previous section, this might not be prudent, but a controlled experiment with one of these tools might be worthwhile in the discovery phase, to see what the fuss is about, so to speak. The other end of the spectrum is building a foundation model for your firm's personal use. There may be a few well-funded conglomerates for whom this is a viable option, but it's out of reach for most enterprises.

There are two ways to leverage generative AI that most enterprises would find manageable: by accessing a foundation model through an intermediary or fine-tuning a model on their own data. The former involves working with one of the major players, like Google or Amazon, who can help a business create a walled garden for generative AI access. There are various configurations, but generally these solutions offer cloud-based access to the model that keeps company data private both from the outside world and the provider itself. While there would be less to worry about the application leaking sensitive data or using proprietary info to train the foundation model, all the main commercial AI products still come with the transparency, model explainability, and bias issues outlined above.

The last option is using a model fine-tuned on transparent data sources or even fine-tuning a transparent foundation model yourself. There are downsized LLMs available that have been condensed from the full-sized parent model, like OpenAI's Codex, which has been trimmed and honed for programming tasks. There are a fair few fine-tuned models available open source and the enterprise would need to run them on its own hardware. They can be used as-is or trained further on proprietary data. This is one of the options we've provided for our customers as we've integrated generative AI into Altair® RapidMiner®, our data analytics and AI platform. We've built an extension that will both simplify users' access to intermediary models like ChatGPT and allow them to build LLMs for new or proprietary use cases – all without needing to write any code. Users will be able to access more than 300,000 Huggingface models with one click. Furthermore, users can fine-tune models with hundreds of billions of parameters with ease, competing with some of the biggest models on the market.

A benefit of these tailored models is that they can be tuned with only a small amount of labeled data, and they learn quicker thanks to their smaller size. Data transparency helps mitigate inherent bias, copyright infringement, and the unreliability of outputs. Keeping generative AI to tasks it's well suited for will lessen the environmental impacts of intensive computing, too.

INDUSTRY-SPECIFIC GENERATIVE AI USE CASES

In addition to the information we've gathered for everyone interested in generative AI, we've also collected some helpful insights for specific industry groups. There are three industries we've put together additional information for:

- Banking, financial services, and insurance (BFSI)
- Manufacturing, including high-tech manufacturing and biomedical/scientific research
- Aerospace and defense

Banking, Financial Services, and Insurance

The BFSI industry is already making good use of machine learning technology – and there's still room to grow in that regard. Machine learning models are already in place at many banks and insurance companies seeking to accelerate underwriting decisions and loan applications, as well as monitoring for financial crime and fraud. These are great uses for discriminative machine learning, tasks that are well suited to the technology's superhuman skill at finding patterns and outliers.

Consultants and thought leaders in the BFSI sector have expressed interest in using generative AI to automate data gathering and report generation for compliance purposes. Another use case BFSI folks are excited about is extracting data from corporate disclosures, quarterly earnings reports, and other unstructured data and generating a summary report on the findings. In each of these cases, generative AI can do this task, but it would likely be more cost-effective to use a discriminative model to do the data gathering, extraction, and synthesis, and saving the report-writing process for the LLM.

Manufacturing

Manufacturing and heavy industry can sometimes be on the cautious side; the large capital investitures these organizations often have to make fosters risk aversion and a measured approach to adopting new technologies. But there are job functions and analytical tasks where discriminative AI models are already proving their worth. A lot is being left on the table here, metaphorically speaking, in terms of capacity for automation. There are several areas where machine learning is currently being deployed, with room for even greater adoption, including:

- Scrap reduction
- Downtime reduction
- Production scheduling
- Identifying and solving process bottlenecks
- Sales price optimization
- Demand prediction and inventory planning
- Warranty claim status and documentation analysis
- Anomaly detection and quality control

Looking into the future, a combination of connected Internet of Things (IoT) devices, computer vision (which relies on neural networks), and generative AI could create some helpful new solutions for product designers and maintenance technicians. Designers and engineers are already benefiting from augmented generation of schematics and advanced model simulations using the generative design capabilities of the Altair® HyperWorks® products. Machine learning-assisted design accelerates product development by speeding up model design and creation of proofs-of-concept (POCs).

On the shop floor, technicians could have a connected "copilot" device – augmented reality glasses or a touchscreen tool that can access a repository of the user manuals for every machine in the place, paired with maintenance records and information on availability of spare parts and other technicians. The repairperson could use natural language to describe the problem to the AI and the device could query all that assembled data and produce a tailored response to the problem at hand. All of these capabilities already exist in piecemeal form, but a powerful generative AI model could combine them into a super-service option.

There are also options for high-tech manufacturers. Semiconductor makers in particular are exploring generative design of electrical circuits, wafer design and floorplanning, and the design of purpose-built integrated circuits. Discriminative models are already being used in some factories for automated yield learning in integrated circuit design. Computer vision could do the high-stakes work of inspecting wafers for defects. Reinforcement learning models could be used to continually optimize manufacturing parameters in an industry where margins of error are getting finer and finer.

Many industry leaders in biotechnology, pharmaceuticals, materials engineering, and similar disciplines are investigating the use of generative AI to help develop new materials, molecules, compounds, messenger proteins, and so on. If industry and the copyright offices around the world find a way to make this work, the effect on product development in these fields could be dramatic.

Researchers and scientists who use complicated equipment have expressed interest in using natural language commands to interact with their instruments and sensors rather than manually coding their operation instructions. This would save time and increase accuracy, a great thing in expensive-to-run high-tech labs.

Aerospace and Defense

Enterprises in aerospace engineering and defense manufacturing have special needs and high standards. Security is paramount for essentially every aspect of their business and so they will want to be especially mindful of the privacy risks associated with generative AI models. All data will need to be totally isolated. The good news is that defense contractors and national departments of defense are some of the few entities with the monetary and personnel resources to build their own foundation models.

Generative AI could open up new avenues for training personnel. Using a combination of LLMs and text-to-image generators, flight instructors would be able to spin up new virtual reality training scenarios in simulation, including scripts and lifelike visuals and audio, specific to a location or set of conditions for pilots and crew. Likewise, military leadership could create new scenarios for war games and teaching strategy, or predicting outcomes of conflicts based on historical data.

These groups are also tasked with monitoring cyberoffensives and information warfare in the interest of national security and will need access to the same – or better – tools as their adversaries to stay one step ahead.

SECURE, TRANSPARENT GENERATIVE AI TOOLS WITH ALTAIR RAPIDMINER

The Altair RapidMiner platform offers comprehensive, end-to-end solutions, from data ingestion and modeling to operationalization and visualization. It's designed for many different skill sets, from data scientists and engineers to business analysts and executives, to get more value from data.

We've added generative AI to our solutions to help bring the power of LLMs to our customers in a safe and stable way. Altair RapidMiner makes it easy for customers to access and build generative AI models and has further accelerated workflow design with generative AI.

To learn more, visit altair.com/altair-rapidminer.

Altair is a global leader in computational science and artificial intelligence (AI) that provides software and cloud solutions in simulation, high-performance computing (HPC), data analytics, and AI. Altair enables organizations across all industries to compete more effectively and drive smarter decisions in an increasingly connected world – all while creating a greener, more sustainable future.

For more information, visit www.altair.com

#ONLYFORWARD