**Generative AI**

**An Overview**

Generative AI refers to the technology that can analyze data to discover relationships and predictable sequences, allowing it to create new outputs such as text, images, and sounds. This is achieved through models that can represent and adapt to various forms of data.

**A Brief History of Artificial Intelligence** The history of AI dates back to 1950 when Alan Turing, often referred to as the father of AI, introduced the concept of machines capable of intelligent thought. Since then, AI has evolved significantly, with developments in neural networks and generative AI driven by advances in hardware and algorithms.

**Data Annotation**

Annotated data, labeled manually by humans, is crucial for supervised learning. This process is time-consuming and expensive but necessary for training accurate models.

Take for instance that you want to develop a program to single out dogs in images. You must go through the rigorous process of feeding it with multiple labeled pictures of dogs and “non-dogs” to help the model learn what dogs look like. The program will then be able to compare new images with its existing repository to find out whether an image contains a dog in it.

Though the process is repetitive at the beginning, if enough annotated data is fed to the model, it will be able to learn how to identify or classify items in new data automatically without the need of labels. For the process to be successful, high quality annotated data is required. This is why most developers choose to use human resources for the annotation process.

The process might be automatized by using a machine to prepopulate the data, but a human touch and a human eye is preferred for review when the data is nuanced or sensitive. The higher the quality of annotated data fed to the training model, the higher the quality of the output. It is also important to note that most AI algorithms require regular updates to keep up with changes. Some may be updated as often as every day.

**Types of Annotation in Machine Learning**

1. **Text Annotation -** Text annotation is the process of attaching additional information, labels, and definitions to texts. Since written language can convey a lot of underlying information to a reader such as emotions, sentiment, stance, and opinion, in order for a machine to identify that information, we need humans to annotate what exactly it is in the text data that conveys that information. Here are some of the examples of text annotation models that can help automate repetitive tasks. The sentiment annotation, Text classification (Document classification also Product categorization), and Entity annotation.
2. **Image Annotation –** Image annotation is to make object recognizable through AI and ML models. It is the process of adding pre-determined labels to images to guide machines in identifying or blocking images. It gives the computer, vision model information to be able to decipher what is shown on the screen. Depending on the functionality of the machine, the number of labels fed to it can vary nonetheless, the annotations must be accurate to serve as a reliable basis for learning. Here are the different types of image annotation: Bounding boxes, Line annotation, 3D cuboids, Landmark annotation, Image transcription.

*Natural language processing (NLP) solutions such as chat bots, automatic speech recognition, and sentiment analysis programs would not be possible without text annotation. To train NLP algorithms, massive datasets of annotated text are required.*

There are several types of AI models, but the major ones are:

* **Supervised Learning Models:** These models learn from labeled data, inferring a function from a given input to the output.
* **Unsupervised Learning Models:** These learn patterns from input data without labeled responses.
* **Reinforcement Learning Models:** These are trained to make sequences of decisions. The goal is to find the best possible way to behave in a given situation.

**AI Algorithm**

An algorithm is the heart of any AI model. For beginners, it's important to understand the various types of AI algorithms and which ones are suitable for different kinds of problems.

Consider the following:

* **Overview of Common AI Algorithms:** Familiarize yourself with algorithms like linear regression, decision trees, and neural networks.
* **Selecting the Appropriate Algorithm:** This depends on the nature of the data and the problem. Some algorithms may be better for structured data, while others excel with unstructured data.

**Training AI Models**

Generative AI models require vast amounts of data and significant computational resources to develop representations that are generalizable and adaptable. By incorporating institutional knowledge, existing models can be fine-tuned to excel in specific use cases.

Foundation models are created by training deep learning algorithms on massive volumes of raw, unstructured data. This involves processing terabytes of data from the internet or other large sources. During training, the algorithm evaluates millions of feedback loops, predicting elements in images or text and adjusting itself to minimize errors.

**Considering the following questions:**

* Who defines responsible use of generative AI, especially as cultural norms evolve and social engineering approaches vary across geographies? Who ensures compliance? What are the consequences for irresponsible use? ​
* In the event something goes wrong, how can individuals take action?
* How do users give and remove consent (opt in or opt out)? What can be learned from the privacy debate?
* Will using generative AI help or hurt trust in your organization — and institutions overall?
* How can we ensure that content creators and owners keep control of their IP and are compensated fairly? What should new economic models look like?
* Who will ensure proper functioning throughout the entire life cycle, and how will they do so? Do boards need an AI ethics lead, for example?

**Risks of Generative AI**

* **Lack of transparency.** Generative AI and ChatGPT models are unpredictable, and not even the companies behind them always understand everything about how they work.
* **Accuracy.** Generative AI systems sometimes produce inaccurate and fabricated answers. Assess all outputs for accuracy, appropriateness and actual usefulness before relying on or publicly distributing information.
* **Bias.** You need policies or controls in place to detect biased outputs and deal with them in a manner consistent with company policy and any relevant legal requirements.
* **Intellectual property (IP) and copyright.** There are currently no verifiable data governance and protection assurances regarding confidential enterprise information. Users should assume that any data or queries they enter into the ChatGPT and its competitors will become public information, and we advise enterprises to put in place controls to avoid inadvertently exposing IP.
* **Cybersecurity and fraud.** Enterprises must prepare for malicious actors’ use of generative AI systems for cyber and fraud attacks, such as those that use deep fakes for social engineering of personnel, and ensure mitigating controls are put in place. Confer with your cyber-insurance provider to verify the degree to which your existing policy covers AI-related breaches.
* **Sustainability.** Generative AI uses significant amounts of electricity. Choose vendors that reduce power consumption and leverage high-quality renewable energy to mitigate the impact on your sustainability goals.
* '**Hallucinations**' are outputs that are nonsensical or inaccurate but seem plausible. For instance, a lawyer once used a generative AI tool for research, only to find that it had produced entirely fictional cases. To mitigate such issues, developers implement guardrails—preventive measures that restrict the model's output. Continuous evaluation and tuning are essential to minimize inaccuracies.

**Ethical Implementation** - for ethical implementation, the data used should be accurate, relevant, and free from biases, hate speech, and other toxic elements. Additionally, models should be trained on legally accessed, high-quality data.