-Traditional AI, often referred to as machine learning (ML), has primarily focused on analytic tasks like classification and prediction.

-Generative AI (gen AI) goes a step further with its ability to create new, original content.

-Gen AI is a type of artificial intelligence that uses neural networks and deep learning algorithms to identify patterns within existing data as a basis for generating original content.

-By learning patterns from large volumes of data, gen AI algorithms synthesize knowledge to create original text, images, audio, video, and other forms of output.

-ML, which involves training computer algorithms to learn patterns and make predictions based on data, emerged in the 1980s.

-These software systems use interconnected nodes (neurons) to process information.

-deep learning revolutionized the AI landscape with its capability to handle large amounts of data and execute complex tasks.

-deep learning employs multiple layers of interconnected neurons, allowing for more sophisticated learning and representation of data.

-This breakthrough led to significant advancements in computer vision, speech recognition, and natural language processing (NLP), launching the era of general-purpose AI bots such as Siri and Alexa.

-Convolutional neural networks (CNNs) proved themselves to be particularly successful at computer vision tasks,

-recurrent neural networks (RNNs) excelled in sequential data processing, such as language modeling. These technologies laid the foundation for gen AI.

-A transformer in artificial intelligence (AI) and machine learning (ML) refers to **a neural network architecture that has become fundamental for processing sequential data**,

**Introducing LLMs and foundation models**

-Large language models (LLMs) are advanced AI systems designed to understand the intricacies of human language and to generate intelligent, creative responses when queried.

-Successful LLMs are trained on enormous data sets typically measured in petabytes (a million gigabytes).

-Training data has often been sourced from books, articles, websites, and other text-based sources, mostly in the public domain.

-Using deep learning techniques, these models excel at understanding and generating text similar to human-produced content.

**Transforming the AI landscape**-a key breakthrough came in 2017 when the Google Brain team introduced the transformer architecture, a deep learning model that replaced traditional recurrent and convolutional structures with a new type of architecture that’s particularly effective at understanding and contextualizing language, as well as generating text, images, audio, and computer code.

-Perhaps the best-known example is OpenAI’s ChatGPT, which stands for chatbot generative pretrained transformer

**Accelerating AI functions**

- important factor in the evolution of AI is the advent of accelerated hardware systems known as graphics processing units (GPUs).

- central processing units (CPUs) are designed for general-purpose computing tasks.

-GPUs, initially developed for graphics rendering, are specialized processors that have proven to be adept at ML tasks due to their unique architecture.

-Transformers use GPUs to process multiple threads of information, leading to faster training of AI models that effectively handle not just text but also images, audio, and video content.

**The Role of Data in AI Projects**

**-**gen AI applications that have been built on public data can’t realize their full potential in the enterprise until they’re coupled with enterprise data stores.

-Most organizations store massive amounts of data, both on-premises and in the cloud.

-Many of these businesses have data science practices that leverage structured data for traditional analytics, such as forecasting.

-To maximize the value of gen AI, these companies need to open up to the vast world of unstructured and semistructured data as well.

-80 to 90 percent of data is unstructured — locked away in text, audio, social media, and other sources. For enterprises that figure out how to use this data, it can provide a competitive advantage, especially in the era of gen AI.

-To amass a complete data set, consider not only your internal first-party data, but also second-party data from partners and suppliers, and third-party data from a service provider or data marketplace.

• First-party data is internal data produced via everyday business

interactions with customers and prospects.

• Second-party data is produced by or in collaboration with trusted

partners, such as product inventory data shared with an e-commerce

or retail sales channel.

• Third-party data can be acquired from external sources to enrich

internal data sets. Common examples include manufacturing supply chain data and financial market data.

**Explaining the Importance of Generative AI to the Enterprise**

-**LLMs for content generation** Gen AI can streamline content creation by generating various types of media, including text, sound, and images.

- One popular use of this technology in the enterprise is to develop chatbots that engage in conversational interactions with business users, helping them obtain accurate answers to their questions.

-By harnessing private data such as customer transaction histories and customer service records, these systems can even deliver personalized content to target audiences while maintaining data security.

-LLMs are also adept at analyzing documents, summarizing unstructured text, and converting unstructured text into structured table formats.

-**LLMs as logical reasoning engines** Within the field of AI, natural language understanding (NLU) focuses on comprehending the intricate meaning in human communication. LLMs can unravel the underlying meaning in textual data, such as product reviews, social media posts, and customer surveys.

-This makes them valuable for sentiment analysis and other complex reasoning tasks that involve extracting meaningful insights from text and providing a deeper understanding of human language.

-**LLMs as translation engines** By leveraging this understanding, LLMs can accurately convert text from one language to another, ensuring effective and reliable translation.

-**LLMs for text retrieval, summarization, and search .** They can search through large databases or the Internet in general to locate relevant information based on user-defined queries.

-For example, a tech company might use an LLM to optimize content for search engines by suggesting relevant keywords, giving visibility into common search queries associated with the topic, and ensuring crawlability.

**Pretrained models**

-There’s a rapidly growing market for creating and customizing gen AI foundation models in many different industries and domains.

-This has given rise to a surge of LLMs that have been pretrained on data sets with millions or even billions of records, allowing them to accomplish specific tasks.

-Nvidia debuts new AI tools for biomolecular research and text processing,” MegaMolBART  
(part of the NVIDIA BioNeMo service and framework).

-can understand the language of chemistry and learn the relationships between atoms in real-world molecules, giving researchers a powerful tool for faster drug discovery.

-Pharmaceutical companies can fine-tune these foundation models using their own proprietary data.

-Training these commercial foundation models is an immense effort that costs tens of millions of dollars.

-Thousands of opensource models are available on public sites such as GitHub and Hugging Face. Developers can use the pretrained AI models as a foundation for creating custom AI apps.

**Security versus ease of use**

-All logical reasoning engines need data to function.

-they become even more powerful and relevant when they’re trained with enterprise data.

-Gen AI systems learn from data; the more data they can access, the more capable they become.

-How can you achieve this without squelching innovation? You start by unifying data in a comprehensive repository that multiple-workgroups can access easily and securely. This allows you to centralize data governance and democratize access to gen AI initiatives across your organization while minimizing complexity and optimizing costs.

**Managing Gen AI Projects with a Cloud Data Platform**

**-**A cloud data platform is a specialized cloud service optimized for storing, analyzing, and sharing large and diverse volumes of data.

-It unifies data security and data governance activities by ensuring that all users leverage a single copy of data.

-A cloud data platform extends your AI horizons by allowing you to store your first-party data and leverage a network of data from second- and third-party data providers as well.

-Data is your core differentiator in the age of gen AI.

-to consolidate disparate sources into a cloud data platform that provides strong security and governance for data and the models customized with that data.