



INTERNATIONAL
INSTITUTE FOR
ANALYTICS™

DATA AND ANALYTICS GLOSSARY
RESEARCH & ADVISORY NETWORK



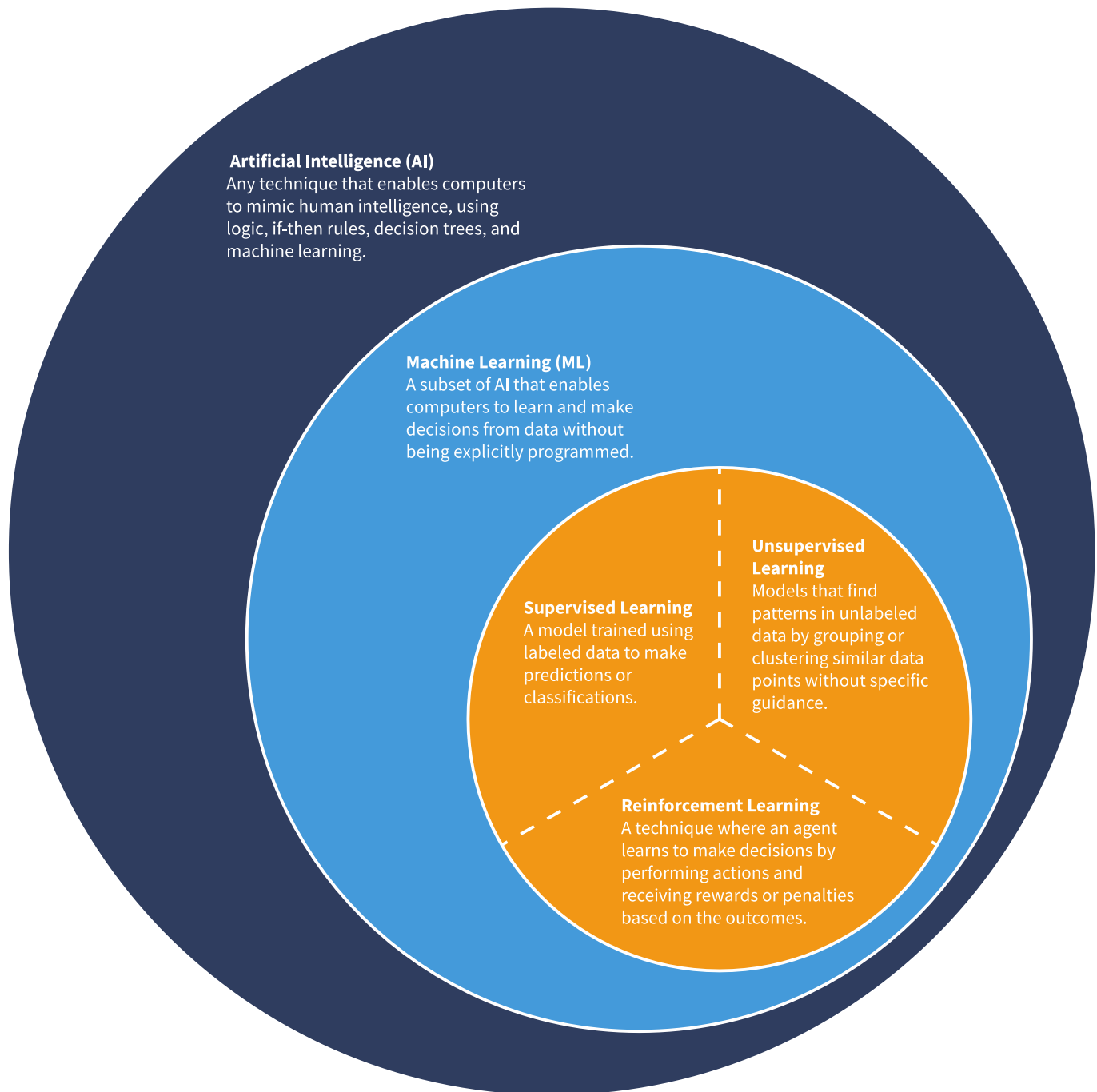
CLIENT ONLY

AI Hierarchy

The purpose of this hierarchy and glossary of terms is to provide definitional clarity and context to key terms and concepts related to artificial intelligence. These definitions have been created to align with IIA's perspective and our experience with clients. It's important to note that the distinctions between AI techniques are not always clear. Many techniques and concepts often overlap multiple categories, and do not fit within a strict hierarchy.

Artificial Intelligence

The majority of what we will discuss in this resource fits under the broad heading of artificial intelligence, commonly defined as “any technique that enables computers to mimic human intelligence.” Machine learning underpins the majority of AI. Within machine learning, the major divisions are supervised learning, unsupervised learning, and reinforcement learning. Understanding how these concepts relate puts you on solid footing for most of the field.



Machine Learning

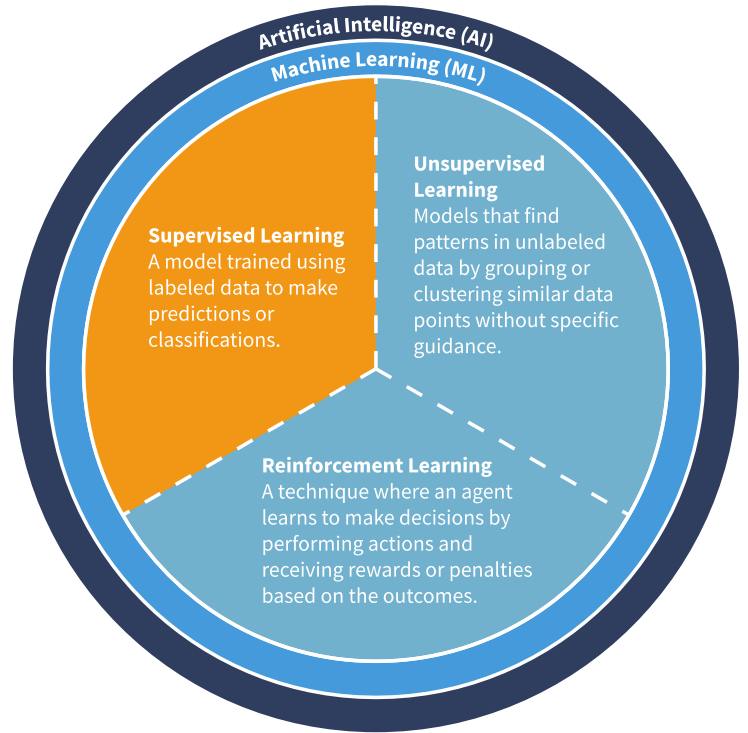
↳ Supervised Learning

MACHINE LEARNING

↳ SUPERVISED LEARNING

↳ REGRESSION

Regression problems in machine learning involve predicting a continuous numerical value or quantity. In regression, the algorithm learns a relationship between input features and a target variable to make predictions. Here are common regression techniques and potential use cases:



	Technique	Definition	Use Case
1	Decision Tree	Decision trees segment data into smaller subsets to make predictions, and they can be used for regression tasks as well.	Analyzing factors influencing customer satisfaction with electronic devices and predicting user ratings based on features like performance, design, and price.
2	Linear Regression	Linear regression models the relationship between a dependent variable and one or more independent variables by fitting a linear equation.	Predicting the number of units of a new electronic product to manufacture based on historical sales data and marketing expenses.
3	Polynomial Regression	Polynomial regression extends linear regression by modeling the relationship between variables using polynomial functions.	Predicting the performance and battery life of smartphones based on various hardware specifications and usage patterns.
4	Support Vector Regression (SVR)	SVR finds the optimal hyperplane to classify data into two or more classes by maximizing the margin between different groups of data points.	Forecasting the price of electronic components based on market demand, historical price trends, and economic indicators.

MACHINE LEARNING

↳ SUPERVISED LEARNING

↳ CLASSIFICATION

Classification problems in machine learning involve categorizing data into predefined classes or labels based on input features. This type of problem is used to assign objects to discrete categories or groups. Here are common classification techniques and potential use cases:

	Technique	Definition	Use Case
1	Naive Bayes	Naive Bayes is a probabilistic classifier often used for text classification based on Bayes' theorem.	Automatically categorizing customer reviews and feedback into predefined topics like “product quality,” “customer service,” or “delivery experience” for analysis.
2	Logistic Regression	Logistic regression is used for binary classification tasks, where the output is a probability that a given input belongs to a particular category.	Identifying defective or non-defective products in electronics manufacturing based on quality control parameters.
3	One-Class SVM	One-Class SVM is used to identify unusual patterns or outliers in data.	Detecting anomalies in sensor data, manufacturing processes, or network traffic to ensure product quality and cybersecurity in the electronics industry.
4	Random Forests	Random forests are an ensemble method used for classifying data into two or more categories, finding the optimal hyperplane to separate classes.	Categorizing customer feedback into multiple sentiment classes, such as “positive,” “neutral,” or “negative,” for comprehensive sentiment analysis.

MACHINE LEARNING

↳ SUPERVISED LEARNING

↳ NATURAL LANGUAGE PROCESSING (NLP)

Supervised NLP involves training algorithms using labeled data. The primary goal of supervised NLP is to develop models that can automatically categorize, analyze, or generate human language based on patterns learned from the labeled training data. Here are common supervised NLP applications and potential use cases:

	Technique	Definition	Use Case
1	Machine Translation	Translating text from one language to another.	Translating user manuals and product documentation to support international customers in using electronics products effectively.
2	Named Entity Recognition (NER)	Identifying and categorizing entities, such as names, dates, and locations, in text.	Extracting product names and specifications from customer feedback for market analysis and product development, such as identifying trends in wearable tech.
3	Sentiment Analysis	Analyzing text to determine the emotional tone, whether positive, negative, or neutral.	Analyzing customer reviews to gauge sentiment and identify areas for product improvement, such as smartphones or smart home devices.
4	Speech Recognition	Converting spoken language into written text.	Transcribing customer service phone calls for quality analysis and identifying common product-related issues in audio interactions.
5	Text Classification	Categorizing text documents into predefined categories.	Categorizing customer support inquiries into topics like troubleshooting, warranty claims, or technical issues for efficient handling.

Machine Learning

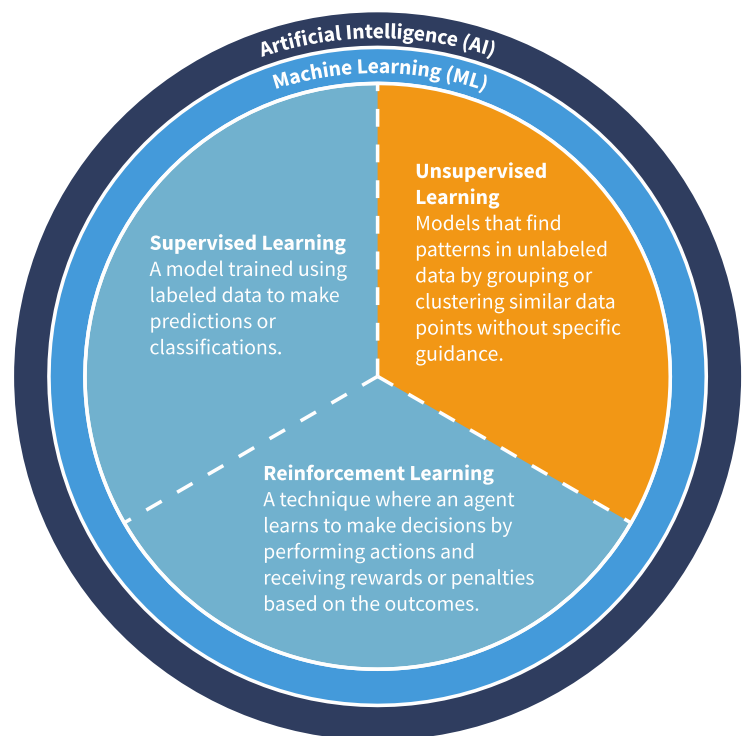
↳ Unsupervised Learning

MACHINE LEARNING

↳ UNSUPERVISED LEARNING

↳ CLUSTERING

Clustering problems in machine learning involve grouping similar data points into clusters or categories based on their intrinsic characteristics or similarities. Clustering is an unsupervised learning technique used to discover patterns or structure within data. Here are common clustering techniques and potential use cases:



	Technique	Definition	Use Case
1	DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	DBSCAN identifies clusters based on data density, forming clusters of varying shapes.	Detecting anomalies or irregularities in sensor data from electronic devices to ensure product quality.
2	Gaussian Mixture Model (GMM)	GMM is a probabilistic model representing data as a mixture of Gaussian distributions, allowing for flexible cluster shapes.	Identifying subgroups within a broader category of electronic products, accommodating various product features and customer preferences.
3	Hierarchical Clustering	Hierarchical clustering creates a tree of nested clusters to reveal data structures at different levels of granularity.	Organizing electronic products into hierarchical categories, helping customers navigate product catalogs efficiently.
4	K-Means Clustering	K-Means is an unsupervised clustering algorithm that partitions data into 'K' distinct clusters based on similarity.	Segmenting customer data into groups for targeted marketing strategies, such as grouping electronic product preferences by user behavior.
5	Self-Organizing Maps (SOM)	SOM is a type of artificial neural network used for clustering data and visualizing high-dimensional data in lower dimensions.	Visualizing customer reviews and feedback on electronic products to identify common patterns and customer preferences.

MACHINE LEARNING

↳ UNSUPERVISED LEARNING

↳ DIMENSION REDUCTION

Dimension reduction in machine learning is the process of reducing the number of input variables (features) used to represent data while preserving the most critical information. This technique is particularly valuable when dealing with high-dimensional data to simplify analysis, improve computational efficiency, and prevent issues like the curse of dimensionality. Here are common dimension reduction techniques and potential use cases:

	Technique	Definition	Use Case
1	Autoencoders	Autoencoders are neural networks used to learn efficient representations of data, effectively reducing dimensionality.	Compressing and representing sensor data from electronic devices to enable faster data analysis and anomaly detection.
2	Feature Selection Techniques (e.g., Recursive Feature Elimination)	Feature selection methods remove irrelevant or redundant features, reducing dimensionality.	Streamlining the analysis of customer feedback and reviews for electronic products by focusing on the most influential features affecting product satisfaction.
3	Linear Discriminant Analysis (LDA)	LDA is used for dimensionality reduction while maximizing the separation between classes in a supervised context.	Improving the classification of electronic components or products based on features like performance, price, and brand reputation.
4	Principal Component Analysis (PCA)	PCA is a linear technique that transforms data into a new coordinate system to reduce dimensionality while preserving as much variance as possible.	Reducing the dimensionality of sensor data from electronic devices to extract essential features and reduce computational complexity.
5	t-Distributed Stochastic Neighbor Embedding (t-SNE)	t-SNE is a nonlinear technique for visualizing high-dimensional data in a lower-dimensional space while preserving neighborhood relationships.	Visualizing and exploring customer behavior and preferences concerning electronic products from high-dimensional user interaction data.

MACHINE LEARNING

↳ UNSUPERVISED LEARNING

↳ NLP

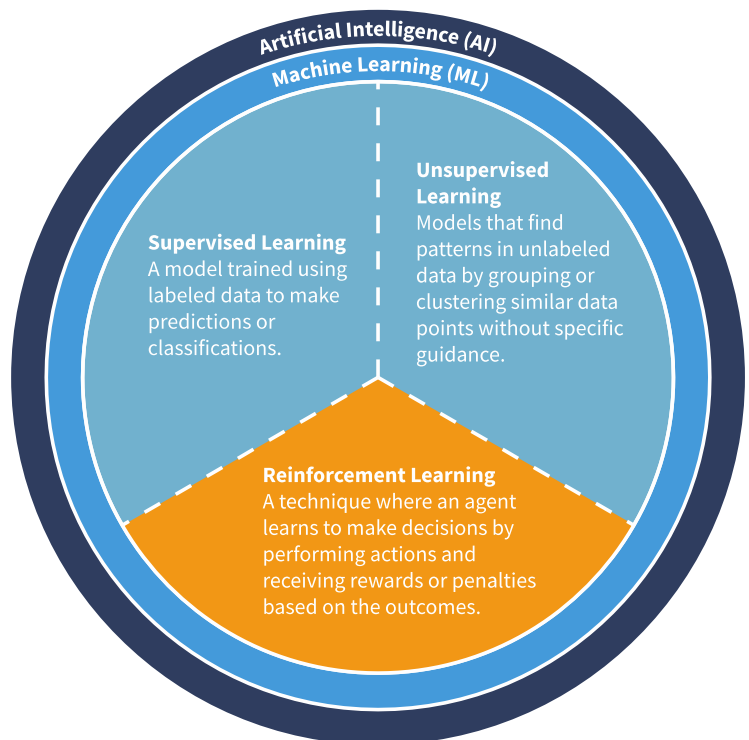
Unsupervised NLP involves algorithms processing and analyzing text data without the need for labeled training data or predefined categories. The primary objective of unsupervised NLP is to discover patterns, structures, and relationships within language data. Here are common unsupervised NLP applications and potential use cases:

	Technique	Definition	Use Case
1	Anomaly Detection	Identifying unusual patterns or outliers in data.	Detecting unusual customer behavior, such as fraudulent product reviews or suspicious interactions on the company's website or support channels.
2	Clustering for Topic Modeling	Grouping and categorizing text data into topics or clusters based on inherent patterns.	Organizing customer reviews and feedback into categories like product features, user experiences, and performance, aiding in product development.
3	Dimensionality Reduction	Reducing data dimensionality while preserving significant variance.	Visualizing customer interactions and reviews to uncover patterns in user behavior and preferences for better-targeted marketing.
4	Text Summarization	Automatically summarizing lengthy text into shorter versions while retaining key information.	Summarizing complex product descriptions or lengthy customer reviews into concise, informative summaries for customers.
5	Word Embeddings	Representing words as dense vectors to capture semantic relationships between them.	Improving sentiment analysis by understanding the context and relationships between words in customer feedback, especially in user reviews of electronic devices.

Machine Learning

↳ Reinforcement Learning

Sequential decision making and exploration vs. exploitation are two fundamental concepts in reinforcement learning. These techniques involve an agent that interacts with an environment over a series of steps or one that decides whether to choose known actions that have yielded rewards in the past (exploitation) or to explore new actions to discover their potential rewards (exploration). Here are common reinforcement learning techniques and potential use cases:



	Technique	Definition	Use Case
1	Actor-Critic Methods	Actor-critic methods combine the strengths of value-based and policy-based approaches, using a critic to estimate value and an actor to determine actions.	Improving the energy efficiency of electronic products by controlling power states and component usage based on real-time demand and user behavior.
2	Deep Q-Network (DQN)	DQN is an extension of Q-learning that leverages deep neural networks to approximate the Q-function, allowing for more complex decision-making in high-dimensional state spaces.	Enhancing the efficiency of inventory management by dynamically adjusting stock levels based on demand forecasts for electronic products.
3	Monte Carlo Tree Search (MCTS)	MCTS is a tree search algorithm used for making optimal decisions in a wide range of applications.	Enhancing the quality control process by optimizing the selection of inspection checkpoints in electronics manufacturing, reducing production costs, and ensuring product quality.
4	Multi-Armed Bandit Algorithms	Multi-armed bandit algorithms involve managing a set of actions with different reward probabilities, allowing the agent to balance exploration and exploitation to maximize long-term rewards.	Improving the efficiency of supply chain management by dynamically selecting suppliers for electronic components based on cost, quality, and historical performance data.
5	Policy Gradient Methods	Policy gradient methods directly learn the optimal policy, making them suitable for problems with high-dimensional action spaces.	Optimizing the marketing strategy for electronic devices by dynamically adjusting advertising budgets and content to maximize sales and customer engagement.
6	Q-Learning	Q-learning is a model-free reinforcement learning technique that learns the quality of actions in a state-action pair and aims to maximize cumulative rewards.	Optimizing the routing of electronic components on the factory floor to reduce production time and minimize energy consumption.
7	Thompson Sampling	Thompson sampling is a probabilistic approach where the agent maintains a probability distribution over action values and samples from this distribution to choose actions.	Maximizing the performance of electronic products by selecting components with uncertain but potentially superior attributes, taking into account both exploration and exploitation in design choices.

MACHINE LEARNING

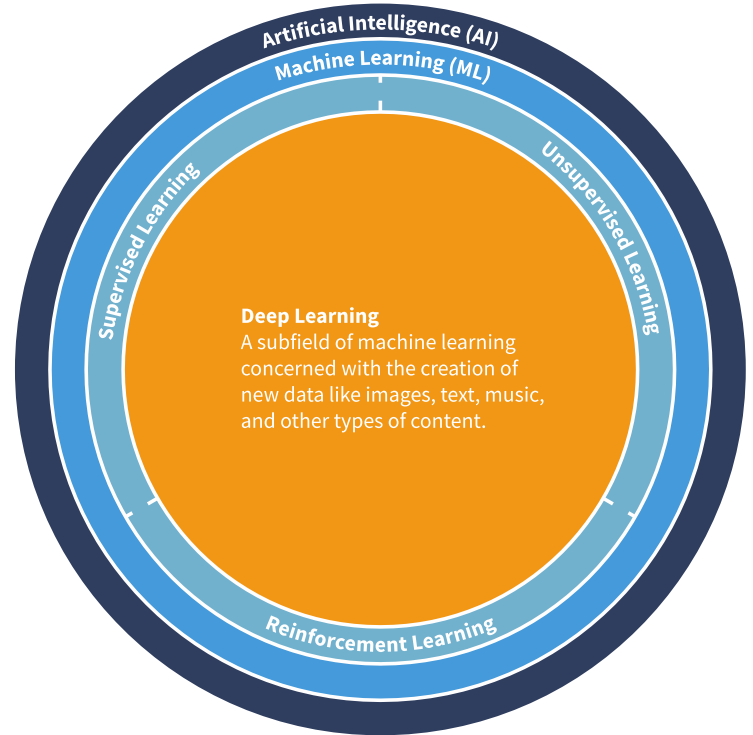
↳ REINFORCEMENT LEARNING

↳ NLP-SPECIFIC APPLICATIONS

	Technique	Definition	Use Case
1	Dialog Systems	Developing intelligent chatbots and virtual assistants to interact with users and improve customer support.	Enhancing customer support by guiding users through troubleshooting, providing accurate responses to inquiries about electronic devices, and assisting in product setup.
2	Language Generation	Generating human-like text and content for various applications.	Creating personalized product descriptions, marketing content, and product recommendations to engage customers in the electronics industry and improve marketing effectiveness.
3	Machine Translation Enhancement	Fine-tuning machine translation models to improve translation quality.	Enhancing the accuracy of translating user manuals, technical documents, and product specifications for global customers who use electronic devices.
4	Automated Content Creation	Automatically generating content for marketing and promotional purposes.	Creating advertising copy, promotional materials, and product descriptions to engage customers in the consumer electronics industry and promote products effectively.
5	Search Engines	Enhancing search engine functionality for improved search results.	Improving the electronic product search experience for customers on the company's website, ensuring they find the products and information they need efficiently.

Deep Learning

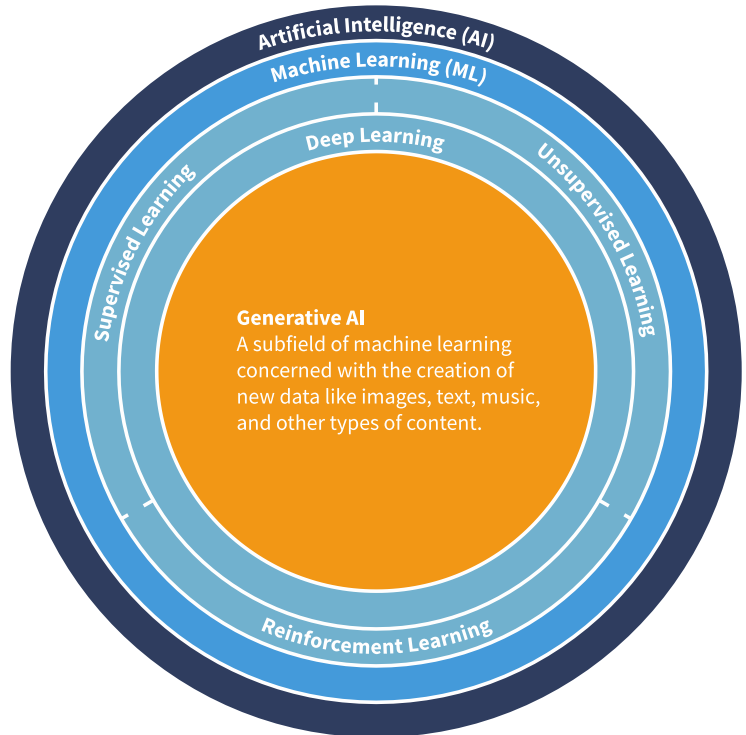
Deep learning utilizes a variety of ML techniques to gain insights from visual data, understand customer sentiment, and optimize operations, to name a few business applications. Deep learning's capacity to handle complex data structures and patterns makes it a powerful tool in various business contexts. Here's a few common techniques and potential use cases:



	Technique	Definition	Use Case
1	Convolutional Neural Networks (CNNs)	CNNs are deep learning models specifically designed for image and spatial data analysis, making them ideal for tasks like image recognition and object detection.	Enhancing quality control in electronics manufacturing by automating the identification of defects or inconsistencies in electronic components.
2	Long Short-Term Memory (LSTM) Networks	LSTMs are a specialized type of RNN designed to capture long-range dependencies in sequential data, making them effective for applications with lengthy context.	Improving customer support for electronic products through the analysis of chatbot interactions, enabling better responses to customer inquiries.
3	Recurrent Neural Networks (RNNs)	RNNs are used for sequential data processing, making them suitable for time series analysis and natural language processing.	Analyzing and predicting equipment maintenance needs based on historical sensor data from electronic devices.

Generative AI

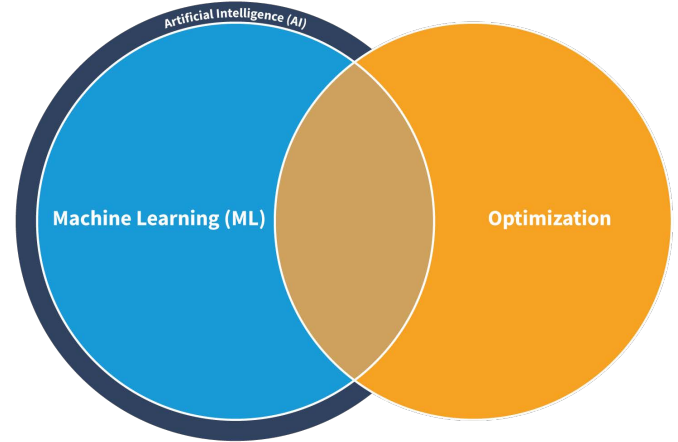
The essence of generative AI is to teach a model to learn the underlying patterns and structures within data so that it can generate new data points that adhere to these patterns. Many generative models leverage deep learning architectures, particularly deep neural networks with multiple layers. Deep learning has proven to be highly effective in capturing complex patterns and features in data, which is crucial for generative tasks. Here are several well-known neural networks used in generative AI:



	Neural Network	Description
1	Variational Autoencoders (VAEs)	VAEs are neural networks that aim to learn representations of data and are useful in applications such as generating synthetic sensor data for testing the quality control systems in electronics manufacturing.
2	Generative Adversarial Networks (GANs)	GANs are a class of neural networks used for generating data and can be applied in electronics design and prototyping, allowing engineers to simulate and evaluate devices before physical production.
3	Recurrent Neural Networks (RNNs)	RNNs are designed for sequential data tasks and are relevant in predicting failures in electronic components over time, enabling proactive maintenance and enhancing product reliability.
4	Long Short-Term Memory (LSTM) Networks	LSTMs excel in capturing long-range dependencies, making them valuable for predicting customer demand for electronic products, aiding in inventory management and supply chain optimization.

Optimization

The field of optimization shares many similarities with AI and ML but is often considered a distinct field. With optimization you are finding the best possible solution among a set of feasible options. It involves using algorithms and mathematical techniques to iteratively refine solutions until an optimal or near-optimal outcome is achieved. Here's a short-list of programming techniques and potential use cases:



	Technique	Definition	Use Case
1	Integer Programming	Integer programming is used to optimize processes with discrete decision variables, making it suitable for problems with integer constraints.	Optimizing production scheduling with discrete time slots for electronic manufacturing.
2	Linear Programming	Linear programming optimizes resource allocation in manufacturing processes by balancing production output with budget constraints.	Efficiently allocate resources in an electronics manufacturing plant, maximizing production, and minimizing costs.
3	Nonlinear Programming	Nonlinear programming optimizes processes involving non-linear objective functions and constraints.	Optimizing complex electronic circuit designs with non-linear relationships between components.
4	Constraint Programming	Constraint programming optimizes processes with logical constraints, making it valuable for problems with intricate rules and dependencies.	Optimizing the configuration of complex electronic systems with numerous interrelated constraints.
5	Dynamic Programming	Dynamic programming is used to optimize processes within manufacturing by intelligently managing the movement of components to enhance efficiency.	Optimize the movement of electronic components in manufacturing for reduced assembly time and improved efficiency.
6	Stochastic Programming	Stochastic programming is an approach used to manage supply chain and inventory challenges while accounting for unpredictable demand fluctuations and supply chain disruptions.	Plan supply chain and inventory management for electronics, ensuring timely production and delivery amid demand fluctuations and supply chain risks.

IIANALYTICS.COM

Copyright © 2024 International Institute for Analytics. Proprietary to subscribers. IIA research is intended for IIA members only and should not be distributed without permission from IIA. All inquiries should be directed to membership@iianalytics.com.