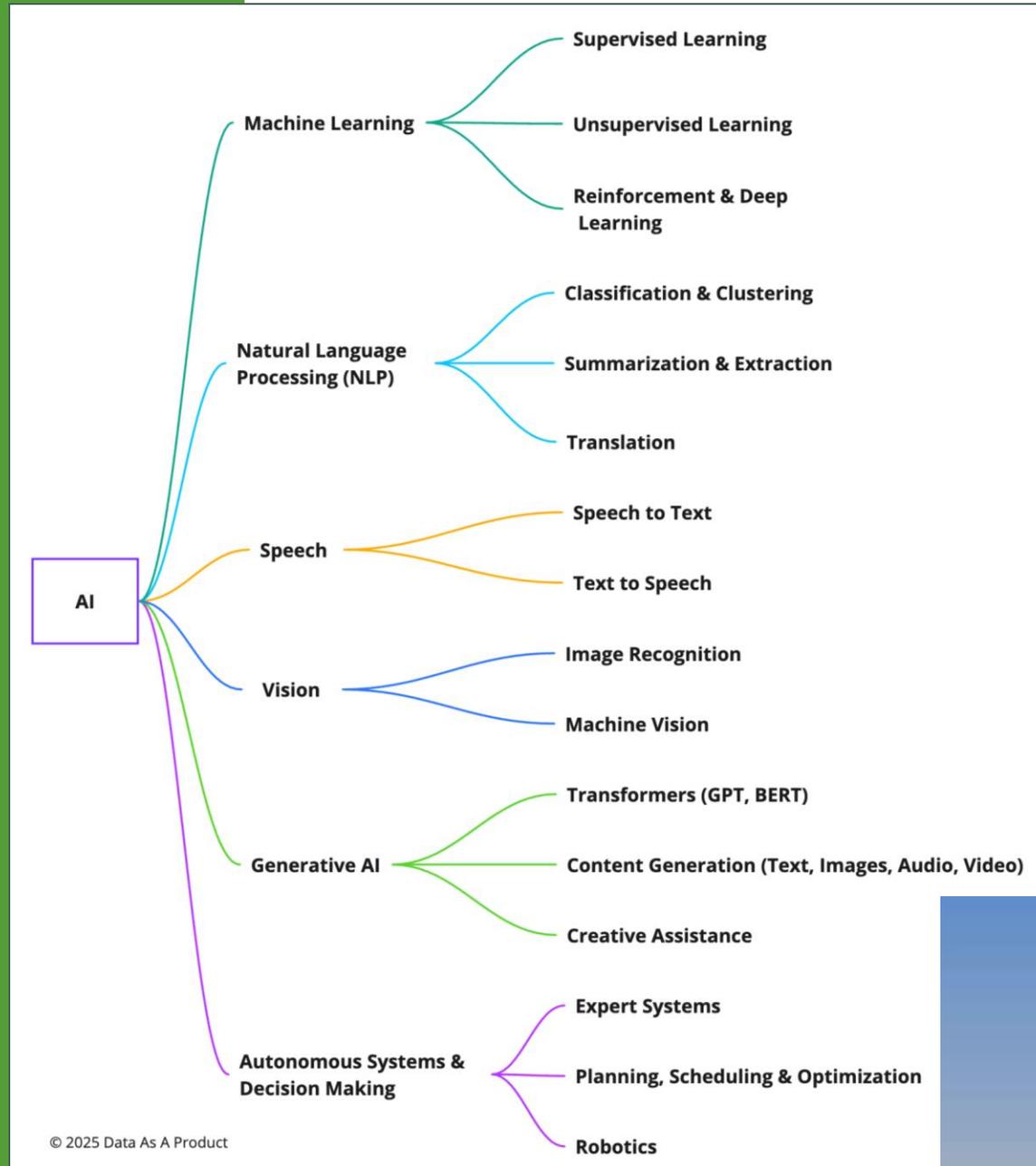


Foundations of AI Products

Module 3

Key Takeaways:

- Understand how AI works, including the types of machine learning, primary algorithms, and the process of turning data into actionable insights.
- Learn the foundations of Machine Learning and Generative AI, and how to apply these concepts to manage AI Products.



Overview of AI Branches

Machine Learning: The backbone of most AI applications, machine learning focuses on building models that learn from data. This branch powers predictive analytics, recommendation engines, and more.

Natural Language Processing (NLP): NLP enables machines to understand, interpret, and respond to human language. From chatbots to sentiment analysis, this branch plays a crucial role in conversational AI.

Vision: AI vision systems process and analyze visual data, such as images and videos. Applications include object detection, facial recognition, and autonomous vehicles.

Speech: This branch focuses on recognizing and synthesizing speech. Voice assistants and speech-to-text tools are examples of this technology in action.

Generative AI: One of the most exciting and transformative branches, Generative AI uses techniques like transformers to create new content, from text and images to code.

Autonomous Systems and Decision-Making: This branch involves systems that can make decisions and act autonomously, such as robotics and supply chain optimization.



Machine Learning Basics

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
What It Is	The model is trained on labeled data. Each data point comes with a known outcome, and the goal is to predict outcomes for new data.	This type of learning involves training models on unlabeled data to find hidden patterns or groupings.	Trains models through a system of rewards and penalties, enabling the model to learn optimal actions in a given environment.
Common Algorithms	Regression, classification, support vector machines (SVM).	Clustering (e.g., k-means), association rules, principal component analysis (PCA).	Q-learning, deep Q-networks (DQN), policy gradients
Applications	Fraud detection, email spam filtering, and predicting customer churn.	Customer segmentation, anomaly detection, and recommendation systems.	Robotics, gaming (e.g., AlphaGo), and autonomous vehicles.



Generative AI Basics

How It Works:

Generative AI models learn patterns in the data they are trained on and use that knowledge to produce original outputs. For example, a text-based model like GPT can generate coherent articles, while an image model like DALL-E can create realistic artwork from textual descriptions.

Key Characteristics:

- *Creativity*: Generative AI mimics human creativity by generating novel outputs.
- *Context Awareness*: These models use contextual understanding to produce relevant and coherent results.
- *Adaptability*: Generative models can be fine-tuned for specific use cases, such as creating product descriptions or designing marketing materials.

Applications:

- *Text Generation*: Tools like ChatGPT can generate human-like text for content creation, customer support, and more.
- *Image and Video Creation*: Models like Stable Diffusion or DALL-E create visuals for design, advertising, and entertainment.
- *Code Generation*: AI models like GitHub Copilot assist developers by generating code snippets based on prompts.



Transformer Architecture and Techniques

The screenshot shows the top of a Google Research blog post. The navigation bar includes 'Google Research', 'Who we are', 'Research areas', 'Our work', 'Programs & events', 'Careers', and 'Blog'. The main title is 'Transformer: A Novel Neural Network Architecture for Language Understanding', dated August 31, 2017, by Jakob Uszkoreit. A bar chart titled 'English German Translation quality' compares four models: GRU4Rec (RNN), ConvS2S (CNN), SockNet (CNN), and Transformer. The Transformer model shows the highest quality score. Below the chart, there is a 'QUICK LINKS' section with a 'Share' button.

Google Research Who we are Research areas Our work Programs & events Careers Blog

Home > Blog >

Transformer: A Novel Neural Network Architecture for Language Understanding

August 31, 2017 · Posted by Jakob Uszkoreit, Software Engineer, Natural Language Understanding

Model	Quality Score
GRU4Rec (RNN)	~45
ConvS2S (CNN)	~50
SockNet (CNN)	~55
Transformer	~65

Neural networks, in particular [recurrent neural networks](#) (RNNs), are now at the core of the leading approaches to language understanding tasks such as [language modeling](#), [machine translation](#) and [question answering](#). In "[Attention is All You Need](#)", we introduce the Transformer, a novel neural network architecture based on a self-attention mechanism that we believe to be particularly well suited for language understanding.

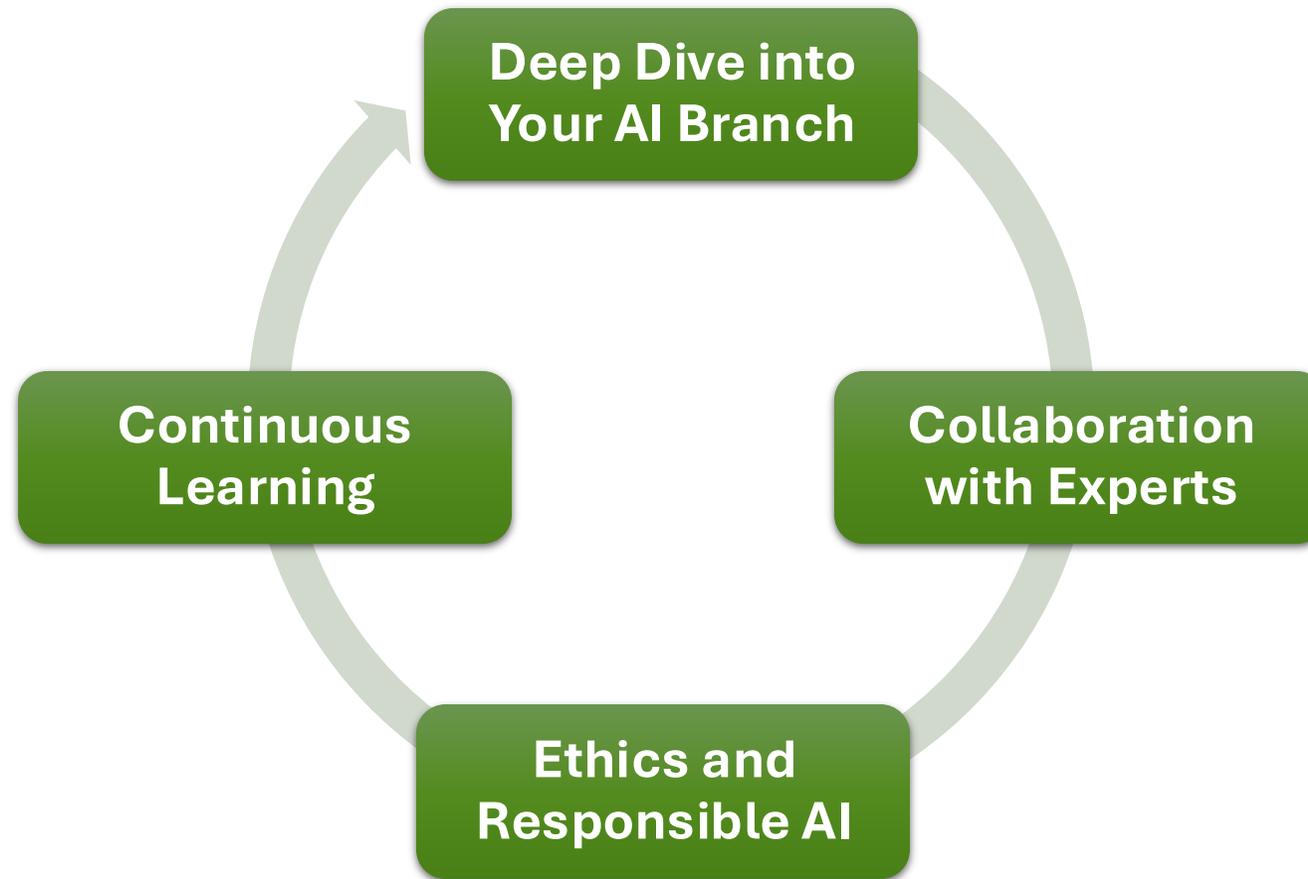
In our paper, we show that the Transformer outperforms both recurrent and convolutional models on academic English to German and English to French translation benchmarks. On top of higher translation quality, the Transformer requires less computation to train and is a much better fit for modern machine learning hardware, speeding up training by up to an order of magnitude.

QUICK LINKS

Share



Building Expertise as an AI Product Manager

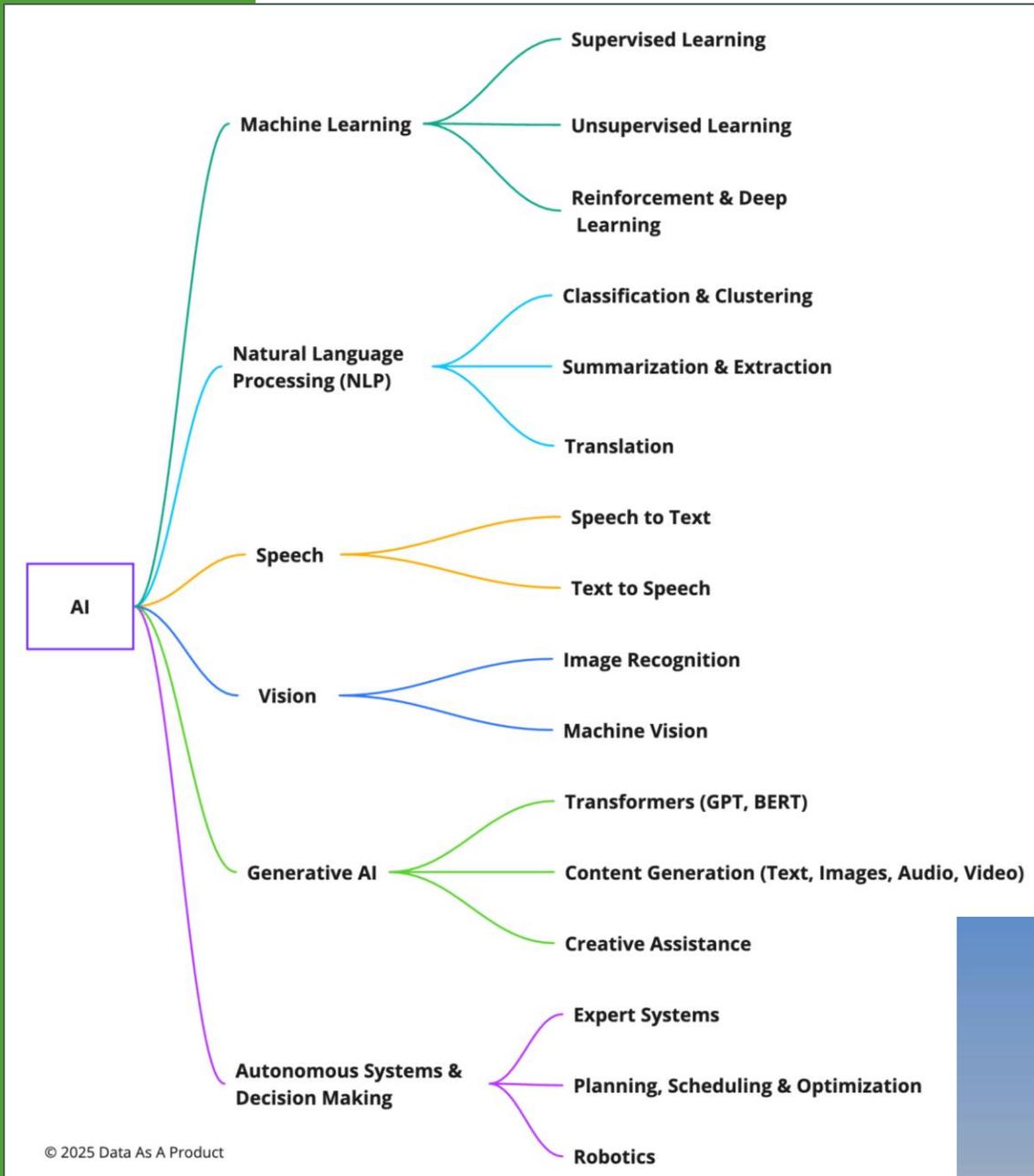


Foundations of AI Products

Module 3

Key Takeaways:

- Understand how AI works, including the types of machine learning, primary algorithms, and the process of turning data into actionable insights.
- Learn the foundations of Machine Learning and Generative AI, and how to apply these concepts to manage AI Products.





Foundations of AI Products

MODULE 3