

Putting AI to Work

3

GPTs and LLMs

Learning Objectives

- Clarify misunderstandings about GPT versions and branding by identifying naming patterns and distinguishing between models and products
- Define large language models (LLMs) and explain how they are trained to understand and generate humanlike language
- Identify and describe the various input types that AI tools can process, such as text, images, and code
- Explore and classify different output formats generated by AI, including summaries, visuals, and structured responses
- Analyze how input and output types are commonly paired in real-world AI use cases to accomplish specific goals

Module 3.1: GPTs

- GPT is a specific OpenAI brand, not a generic term for all AI chatbots.
- Official GPT models:
 - GPT-3
 - GPT-4 from OpenAI (like Oreo being a specific cookie brand)
- Off-brand models use similar technology but different recipes:
 - LLaMA
 - Gemini
 - Claude
- GPT is commonly misused as catch-all term for any transformer-based language model.
- This mislabeling causes confusion about what technology is actually powering AI tools.

Module 3.1: Ethics in Action

- Mislabeling AI models erodes trust and creates confusion for consumers.
- Transparency about model origin, training data, and capabilities is critical.
- Ethical AI use requires honesty in branding and communication.

Module 3.1: Techie Dive

- GPT stands for *generative pre-trained transformer*, a specific architecture.
- OpenAI's GPT models are proprietary, cloud hosted, and trained on enormous datasets.
- Other companies build models using similar transformer techniques but different data.
- The term *GPT* technically refers only to OpenAI's generative pre-trained transformers.

Module 3.1: Business Lens

- Brand recognition matters: Misusing "GPT" can benefit OpenAI's visibility.
- Using the term incorrectly can lead to legal, reputational, and support issues for businesses.
- Marketing must accurately reflect what technology powers AI tools.
- Choosing the right model impacts performance, safety, and integration.

Module 3.2: Large Language Models (LLMs)

- LLMs are AI systems trained to understand and generate humanlike text.
- They are called "large" because they are trained on vast data with billions of parameters.
- They learn patterns from massive text datasets and are not preloaded with facts.
- Training requires filtering biased or harmful content from datasets.
- The energy consumption of training is substantial (comparable to hundreds of homes).

Module 3.2: Ethics in Action

- Biased training data leads to biased model outputs.
- Some training data includes copyrighted content scraped without permission.
- The labeling tasks are often outsourced to low-paid workers in poor conditions.
- Green AI: There is pressure to reduce environmental impact through efficiency.

Module 3.2: Techie Dive

- Training uses unsupervised learning, which is later refined with fine tuning and RLHF.
- Data sources:
 - Common Crawl
 - Wikipedia
 - Books
 - Code repositories
- Model pruning, quantization, and distillation reduce size and energy use.
- LLMs don't think or feel; they mirror human discourse from their training data.

Module 3.2: Business Lens

- Training an LLM costs millions of dollars due to data, compute, and time.
- Few companies develop foundation models; others generally license or customize.
- Some companies create domain-specific models trained on curated data for legal, medical, or financial use.
- The high energy consumption of AI directly affects deployment costs at scale.

Module 3.3: Input Types

- AI systems accept various input types depending on the tool.
- Four primary types:
 - Text inputs include written prompts, questions, and commands for chatbots and writing tools.
 - Image inputs include uploaded photos or drawings for editing, recognition, and generation.
 - Audio inputs include spoken instructions for voice assistants and transcription tools.
 - Video inputs include recorded footage for analysis, surveillance, and education.

Module 3.3: Ethics in Action

- Private photos or voice recordings may compromise privacy.
- Understanding what AI retains and how data is used or stored is critical.
- Sensor data, 3D models, and code files raise specialized ethical concerns.

Module 3.3: Techie Dive

- Multimodal models process multiple input types simultaneously.
- Models like GPT-4 with vision use separate processing paths merged for output.
- Convolutional networks handle images while transformers process text.

Module 3.3: Business Lens

- Multimodal AIs serve broader user needs for customer service and design.
- Tools accepting varied inputs are important for business workflows.
- The choice of input types affects accessibility and inclusion.

Module 3.4: Output Types

AI outputs come in many forms depending on the task and model.

- Text: natural language answers, essays, summaries
- Code: snippets or full programs in various languages
- Image: generated or enhanced visuals
- Audio: AI-generated speech, music, sound effects
- Video: AI-generated clips or animated explainers
- Structured data: charts, spreadsheets, CSV, JSON, XML

Module 3.4: Ethics in Action

- Outputs can include factual errors (hallucinations) or biased phrasing.
- It's the user's responsibility to fact check and interpret outputs critically.
- Generated content raises questions about originality and credit.

Module 3.4: Techie Dive

- Different model types are optimized for different output types.
- Language models use transformer architectures.
- Image generators like Stable Diffusion rely on latent diffusion processes.
- Audio and video tools often blend models across domains.

Module 3.4: Business Lens

- Output format and quality directly impact productivity and branding.
- Businesses must evaluate whether tools produce client-ready outputs.
- Customer trust depends on output accuracy and professionalism.

Module 3.5: Common Input/Output Pairings

AI transforms one kind of data into another based on input/output pairing.

- Text-to-text: writing help, summaries, answers, code suggestions
- Text-to-image: visual storytelling, ads, concept art, diagrams
- Text-to-audio: audiobooks, voice assistants, narration
- Text-to-video: explainer videos, animations, video ads
- Text-to-code: programming help, automation, learning code
- Text-to-structured data: research summaries, tables, comparisons

Module 3.5: Ethics in Action

- Input/output design affects inclusion and fairness.
- Testing across diverse inputs prevents accessibility failures and bias.
- Privacy concerns apply across all transformation types.

Module 3.5: Techie Dive

- Most generative models specialize in one type of input/output.
- Multimodal models are trained to handle multiple types using different architectures.
- Choosing the right tool requires understanding model specializations.

Module 3.5: Business Lens

- Understanding pairings helps teams choose the right tool for the job.
- Training videos need text-to-video and text-to-audio tools.
- Law firms focus on text-to-text summarization and audio-to-text transcription.
- It's important to match AI tools to specific business workflow requirements.

Key Takeaways

- “GPT” is a specific OpenAI brand, although it’s commonly misused as generic term.
- LLMs are trained on massive datasets to predict patterns, not understand meaning.
- Training data quality and biases significantly impact model behavior.
- There are four primary input types: text, image, audio, video.
- There are six primary output types: text, code, image, audio, video, structured data.
- Input/output pairings determine which tools are appropriate for tasks.
- Multimodal models expand possibilities, but specialized tools often perform better.
- Ethical considerations include privacy, bias, accessibility, and consent.
- Business decisions should consider pairings, quality, cost, and compliance.
- Understanding AI fundamentals enables informed tool selection and responsible use.