

Putting AI to Work

2

AI Technologies

Learning Objectives

- Describe the deep learning structures and model types that power generative AI and differentiate between architectures, models, and systems
- Explain how AI systems use interfaces and workflows to integrate models in order to create usable tools for different tasks
- Identify and evaluate common AI-powered products by examining how they are built on underlying systems and models
- Analyze the limitations of generative AI and assess how rapid change affects reliability, usability, and long-term planning
- Examine the risks related to AI accuracy, data privacy, and hallucinations, and propose ways to mitigate them through responsible use

Module 2.1: Architectures and Models

- The AI stack: Architecture → Model → System → Product.
- Architectures are blueprints; models are trained brains.
- Common architectures:
 - Neural networks
 - CNNs
 - RNNs
 - VAEs
 - Diffusion
 - Transformers
- Models can be general purpose or specialized and open or closed.
- Types:
 - LLMs
 - Computer vision
 - Speech/audio
 - Generative image
 - Multimodal
 - Specialized

Module 2.1: Ethics in Action

- AI models can inherit biases from training data.
- Open models promote transparency but may be repurposed for harmful uses.
- Closed models offer control but lack outside oversight.
- Ethical AI requires fairness evaluation, bias testing, and transparency.

Module 2.1: Techie Dive

- Models have three components:
 - Parameters
 - Weights
 - Tokens
- Neural networks are the foundation of modern AI.
- Transformers use "attention" for context understanding.
- CNNs are specialized for vision tasks.
- RNNs handle sequential data.
- Diffusion models transform noise into coherent images.

Module 2.1: Business Lens

- General-purpose models:
 - Writing
 - Summarizing
 - Automation
- Specialized models offer efficiency in niche areas.
- Fine-tuned models adapt to company-specific data.
- Customization is key for law, education, finance, and healthcare.

Module 2.2: AI Systems

- Complete packages include models + tools + interfaces.
- Major systems:
 - OpenAI GPT
 - Google Gemini
 - Meta LLaMA
 - Anthropic Claude
- Systems include APIs, safety layers, and backend infrastructure.
- Systems differ in accuracy, safety, cost, and reliability.

Module 2.2: Ethics in Action

- Systems deliver content and make decisions in high-stakes settings.
- Fairness and bias mitigation must be built in at the system level.
- This is critical in healthcare, hiring, and finance.

Module 2.2: Techie Dive

- Systems include:
 - APIs
 - Training protocols
 - Prompt engineering
 - Evaluation pipelines
- Safety filters screen for harmful content.
- Infrastructure makes systems powerful but complex.

Module 2.2: Business Lens

- Systems are the "black box" delivering results.
- Understanding which system powers a tool helps inform decisions about accuracy, cost, and reliability.
- Consider uptime, support, and integration with existing tools.

Module 2.3: AI Products

- They are user-facing applications built on systems.
- Major products:
 - DALL·E
 - Adobe Firefly
 - ChatGPT
 - Canva
 - Microsoft Copilot
- Products combine multiple AI types for comprehensive functionality.
- DALL·E is better at text integration and conceptual art.
- Firefly is good for photorealistic imagery but struggles with text in images.

Module 2.3: Ethics in Action

- Are AI-generated visuals original or copied?
- Who gets credit for AI-created work?
- Who is responsible when something goes wrong?
- What is AI's impact on employment in creative fields?
- There's a need for transparency and clear usage guidelines.

Module 2.3: Techie Dive

- Products use APIs to connect their front ends with cloud-hosted models.
- The raw output is formatted, filtered, and presented in a user-friendly manner.
- Products cache requests to improve speed and reduce costs.
- Platforms:
 - OpenAI
 - Hugging Face
 - Google Cloud

Module 2.3: Business Lens

- Cut costs through automation.
- Enable mass personalization at scale.
- Raise legal questions about ownership and liability.
- Require employee upskilling.
- Evaluate tools on reliability, support, privacy, integration, and long-term viability.

Module 2.4: Limitations and Rapid Change

- Hallucinations are confident but incorrect outputs.
- Outdated information stemming from training-data cutoffs.
- Different tools give different answers.
- Many industries actively use AI products:
 - Video
 - Music
 - 3D design
 - Education
 - Mental health
 - Gaming
- There are challenges related to building on constantly changing technology.

Module 2.4: Ethics in Action

- Ethical concerns:
 - Transparency: Users don't know what the training data contains or the update dates.
 - Displacement: AI is changing the job landscape in content creation and customer service.
 - Dependence: Over-reliance reduces critical thinking.
 - Accountability: Who is responsible for AI mistakes?
- AI literacy includes knowing when NOT to use tools.

Module 2.4: Techie Dive

- Rapid change drivers:
 - New models
 - API updates
 - Business model shifts
- AI tools are living systems, not static software.
- Real examples:
 - Air Canada chatbot
 - Lawyer fake cases
 - Samsung leak
 - Bard errors
- This creates challenges for businesses building workflows.

Module 2.4: Business Lens

- Risks:
 - Tools change overnight.
 - Free tools become paid.
 - Features are removed.
- Opportunities:
 - The knowledge they provide gives a competitive advantage.
 - It offers businesses new capabilities.
 - Certain tasks will cost less to perform.
- Strategy:
 - Vet for accuracy and privacy.
 - Stay informed.
 - Build flexibly.

Module 2.5: Reliability, Privacy, and Hallucinations

- AI gives the wrong information confidently.
- Privacy risks: What happens to your data?
- Never enter personal identifiers, private company data, or sensitive information.
- Why do hallucinations happen, and how can they be detected?
- Verification of outputs is critical in high-stakes situations.

Module 2.5: Ethics in Action

- Ethical questions:
 - Misinformation: Who is responsible for bad advice?
 - Consent: Are users aware data trains future models?
 - Bias and harm: AI reflects internet biases.
- Vulnerable populations need special considerations.
- Responsible use means knowing what AI CAN'T do.

Module 2.5: Techie Dive

- Risks
 - Reliability: Models are trained on patterns, not truths.
 - Privacy: Free versions store/reuse inputs.
 - Hallucinations: These are most common in long, open-ended tasks.
- Web-search tools (Perplexity, Gemini) offer better reliability, but their outputs still need fact checking.

Module 2.5: Business Lens

- Issues impacting risk, trust, and cost:
 - Risk management: Incorrect outputs can damage reputation and cause liability.
 - Trust building: This stems from transparency about AI use.
 - Cost considerations: Fixing errors is expensive, and human review reduces risk.
- Required policies:
 - Approved tools
 - Data handling
 - Incident response
 - Regular audits

Key Takeaways

- AI is layered: Architecture → Model → System → Product.
- Different architectures excel at different tasks.
- Major systems (GPT, Gemini, LLaMA, Claude) have distinct strengths.
- Products combine models with interfaces for practical use.
- Hallucinations are a fundamental limitation.
- Never enter sensitive data into AI tools.
- Rapid change creates opportunities and challenges.
- Verification, transparency, and human oversight remain essential.