

# AN OVERVIEW OF GENERATIVE AI & LLM

July 3rd, 2025

# OUTLINE

# OUTLINE

- Introduction

# OUTLINE

- Introduction
- What is Generative AI?

# OUTLINE

- Introduction
- What is Generative AI?
- What is LLM?

# OUTLINE

- Introduction
- What is Generative AI?
- What is LLM?
- AI Ethics & Challenges

# OUTLINE

- Introduction
- What is Generative AI?
- What is LLM?
- AI Ethics & Challenges
- Demo

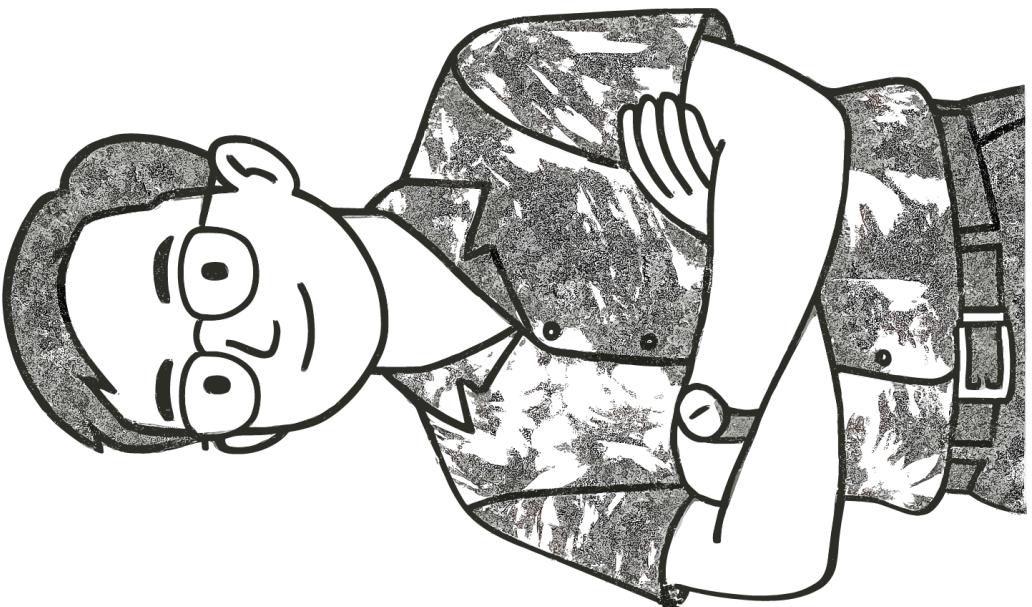
# OUTLINE

- Introduction
- What is Generative AI?
- What is LLM?
- AI Ethics & Challenges
- Demo
- Conclusion

# OUTLINE

- Introduction
- What is Generative AI?
- What is LLM?
- AI Ethics & Challenges
- Demo
- Conclusion
- Q&A

# INTRODUCTION



## EKKY ARMANDI

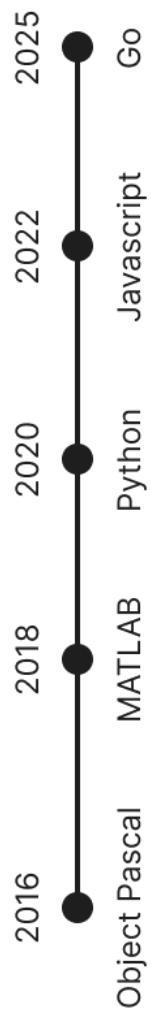
Full-stack Python Developer

Freelancer 🧑 5yr 💼 80+ 🔧 100+

### *Education*

- (2019) S1 Teknik Industri, UAD
- Undergraduate Thesis: Hybrid Genetic Algorithm

### *My programming journey*



*Services:*

- Web Development
- Backend Development
- Data Engineering
- AI Engineering

*Featured Projects:*

- **Trade Arena (2023)** Developing a UI for a stock trading application
- **Real Info (2023-2025)** Developing data scraping tools for collecting 20+ Real Estate websites data in Bali
- **Zixelise (2023-2024)** Developing configuration tools application for Blender 3D Rendering application
- **ArmOrizeIO (2024)** Developing the SaaS backend for a cybersecurity application
- **Bright Data (2024)** Developing scalable data scraping tools and an Telegram bot interface for collecting and managing Real Estate data in Siberia
- **OpsGuide (now)** Developing AI agent and Model Context Protocol (MCP) server

# GENAI

## *What is Generative AI?*

**Generative AI** is a artificial intelligence systems that can create new content such as text, image, code, audio, even video, based on what is trained on





# KEY DRIVERS:

## KEY DRIVERS:

- Creative automation

## KEY DRIVERS:

- Creative automation
- Natural interaction

## KEY DRIVERS:

- Creative automation
- Natural interaction
- Cost reduction

## KEY DRIVERS:

- Creative automation
- Natural interaction
- Cost reduction
- Rapid prototyping

## KEY DRIVERS:

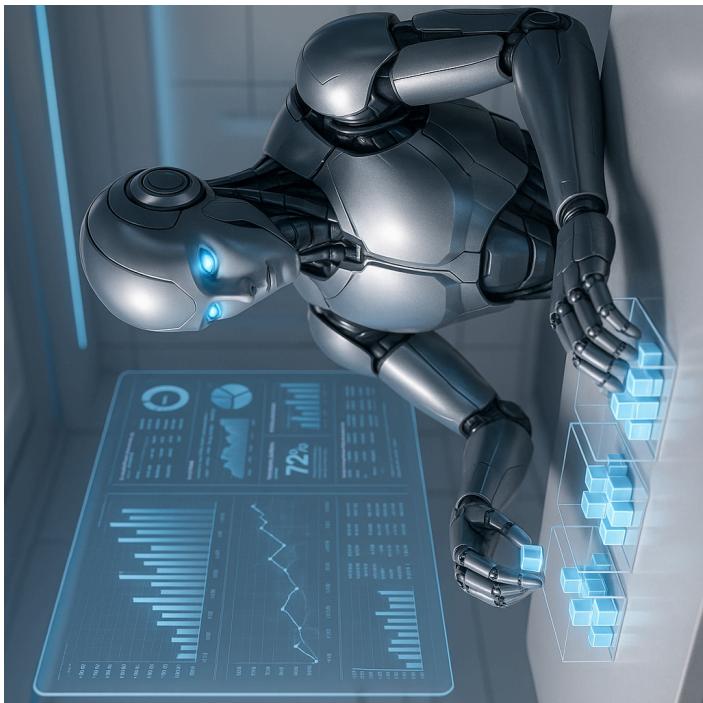
- Creative automation
- Natural interaction
- Cost reduction
- Rapid prototyping
- Cross-industry applicability

# TRADITIONAL AI VS GENERATIVE AI



### Generative AI:

- Creates original content from learned patterns
- Uses models like GPTs, VAEs, GANs, diffusion models
- Output: new text, images, music, code

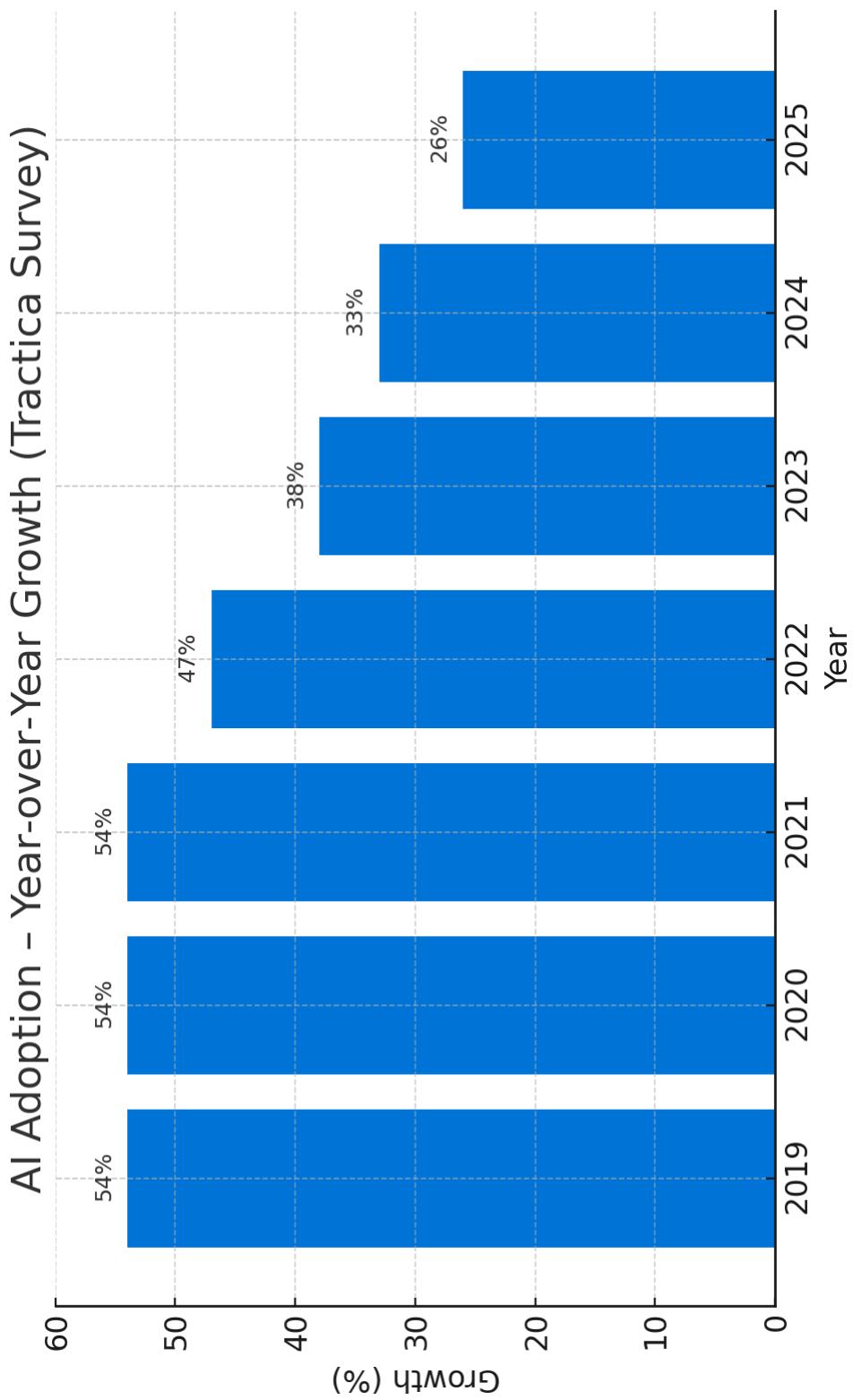


### Traditional AI:

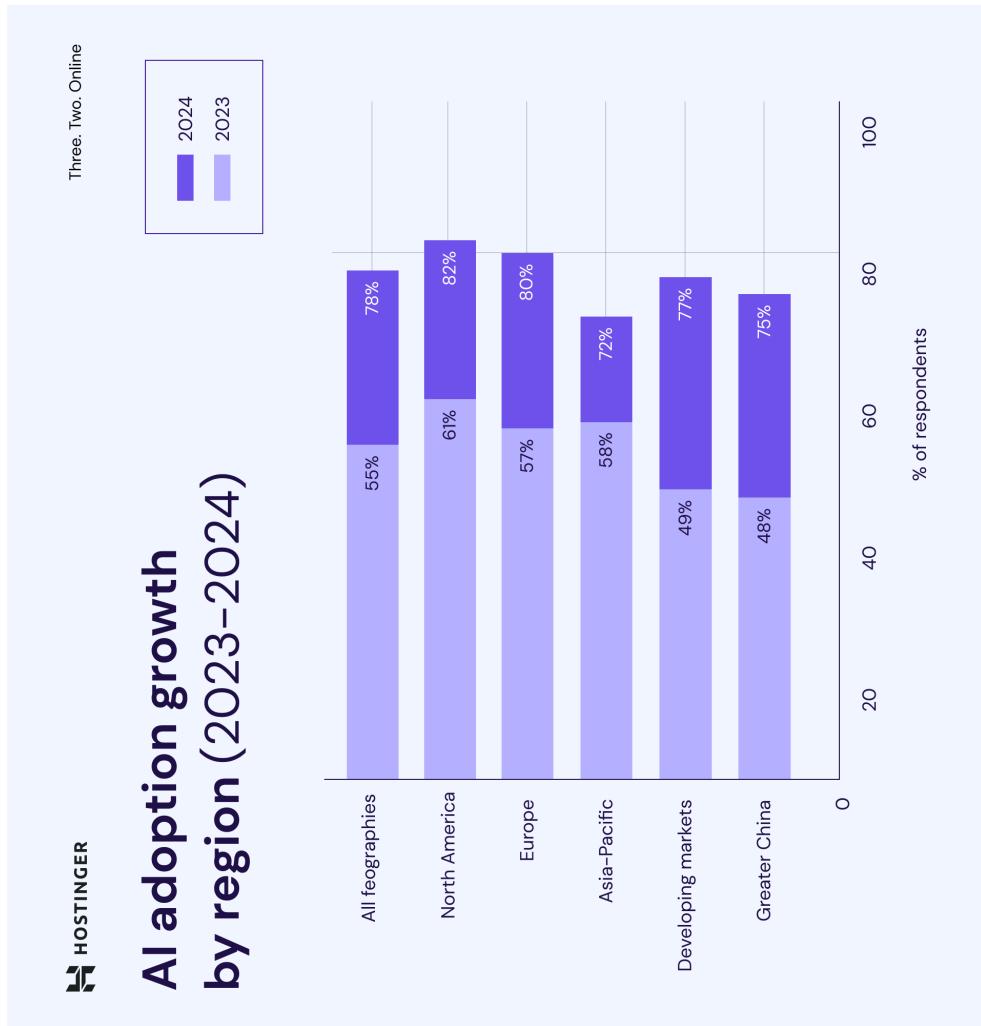
- Classification, prediction, pattern recognition
- Rule-based systems, decision trees, neural networks
- Output: labels, scores, recommendations

# GEN AI IMPLEMENTATION

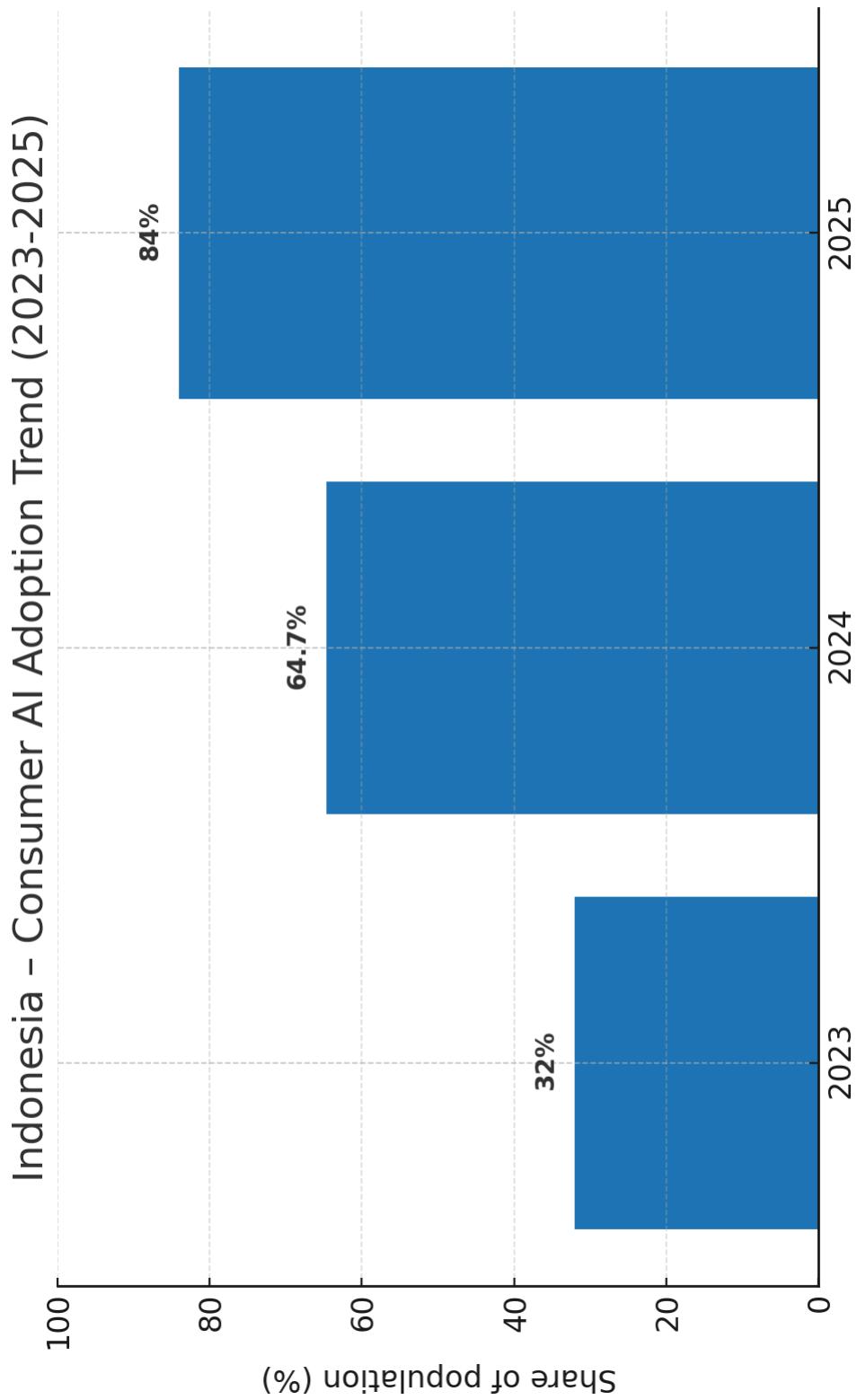
- ChatGPT (text/code/image generation)
- Suno (music generation)
- Elevenlabs (voice generation)
- Midjourney (image/video generation)
- Claude Code (code generation)
- etc.



Source: Tractica Survey



Source: [Hostinger](#)



Source: OpenAI Research

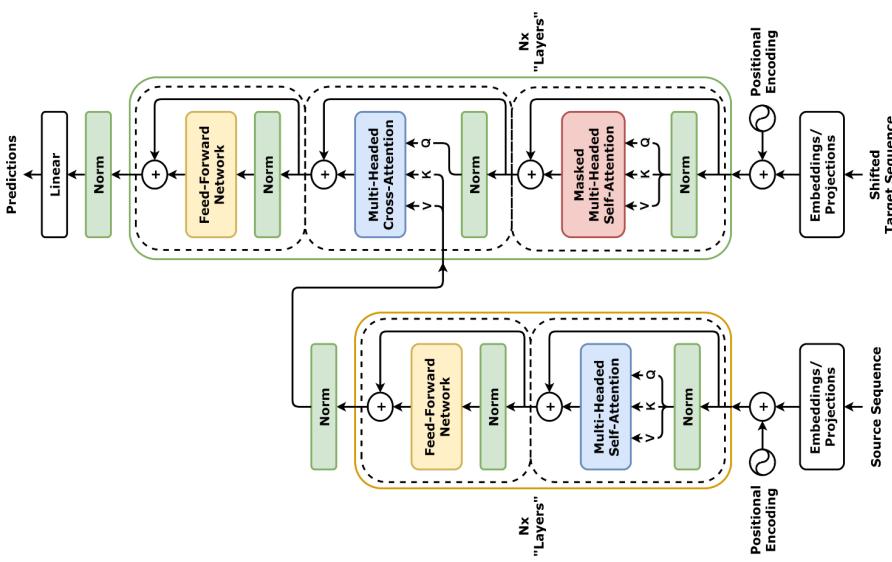
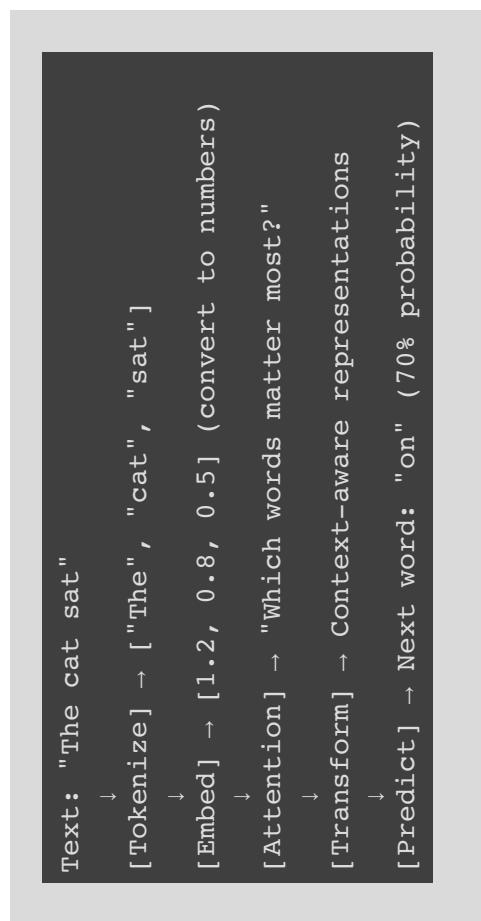
# LARGE LANGUAGE MODEL (LLM)

## *What is LLM?*

**LLM or Large Language Model** is specific Generative AI focus on text generation. LLM is a AI system trained on **massive text dataset** to understand and generate human-like text.



*How it works?*



Source: [Attention is All you Need](#)

# TERMINOLOGY

- **Tokenization** -> Breaking text into smaller units (**tokens**) that the model can process. Example: "ChatGPT is amazing!" → [Chat, G, PT, is, amaz, ing, !]

- **Tokenization** -> Breaking text into smaller units (tokens) that the model can process. Example: "ChatGPT is amazing!" → [Chat, G, PT, is, amaz, ing, !]
- **Prompt** -> The input text/instructions you give to the LLM to get a desired output. Example prompt: "Summarize this article in 3 bullet points"

- **Tokenization** -> Breaking text into smaller units (tokens) that the model can process. Example: "ChatGPT is amazing!" → [Chat, G, PT, is, amaz, ing, !]
- **Prompt** -> The input text/instructions you give to the LLM to get a desired output. Example prompt: "Summarize this article in 3 bullet points"
- **Inference** -> The process of running the model to generate predictions/outputs.

- **Tokenization** -> Breaking text into smaller units (tokens) that the model can process. Example: "ChatGPT is amazing!" → [Chat, G, PT, is, amaz, ing, !]
- **Prompt** -> The input text/instructions you give to the LLM to get a desired output. Example prompt: "Summarize this article in 3 bullet points"
- **Inference** -> The process of running the model to generate predictions/outputs.
- **Zero-shot Learning** -> Model performs a task without any examples in the prompt.

- **Tokenization** -> Breaking text into smaller units (tokens) that the model can process. Example: "ChatGPT is amazing!" → [Chat, G, PT, is, amaz, ing, !]
- **Prompt** -> The input text/instructions you give to the LLM to get a desired output. Example prompt: "Summarize this article in 3 bullet points"
- **Inference** -> The process of running the model to generate predictions/outputs.
- **Zero-shot Learning** -> Model performs a task without any examples in the prompt.
- **Few-shot Learning** -> Model learns from a few examples provided in the prompt.

# MODELS

# HUGGINGFACE.CO

The biggest open-source models repository and one of AI hosting platform.

The screenshot shows a search results page for "Models" on HuggingFace.co. The search bar contains "290,257". Below it are filters: "Filter by name" (selected), "Full-text search", and "↑ Sort: Trending". The results list includes:

- tencent/hunyuan-A13B-Instruct**  
Text Generation · ⚡ 80B · Updated about 6 hours ago · ↓ 5,02k · ⚡ 669
- Menlo/Jan-nano-128K**  
Text Generation · ⚡ 4B · Updated 1 day ago · ↓ 2,68k · ⚡ 178
- baidu/ERNIE-4.5-21B-A3B-PT**  
Text Generation · ⚡ 22B · Updated 2 days ago · ↓ 462 · ⚡ 74
- deepseek-ai/DeepSeek-R1-0528**  
Text Generation · ⚡ 685B · Updated May 29 · ↓ 157k · ⚡ 2,16k
- THU-KEG/LongWriter-Zero-32B**  
Text Generation · ⚡ 33B · Updated 2 days ago · ↓ 632 · ⚡ 81
- MinimaxAI/Minimax-M1-80k**  
Text Generation · ⚡ 456B · Updated 8 days ago · ↓ 24,3k · ⚡ 613
- meta-llama/Llama-3.1-8B-Instruct**  
Text Generation · ⚡ 8B · Updated Sep 26, 2024 · ↓ 5,16M · ⚡ 4,21k
- Menlo/Jan-nano-128k-8guf**  
Text Generation · ⚡ 4B · Updated 1 day ago · ↓ 39,3k · ⚡ 46
- Menlo/Jan-nano**  
Text Generation · ⚡ 4B · Updated 1 day ago · ↓ 34,2k · ⚡ 434

# OPENROUTER.AI

LLM provider and router.

## Models

454 models | Reset Filters

Filter models | Top Weekly | ☰ | ⚙️

### [Anthropic: Claude Sonnet 4](#)

298B tokens

Programming (#1) • Technology (#2) • Health (#3) • Science (#5) • Marketing (#5)

Claude Sonnet 4 significantly enhances the capabilities of its predecessor, Sonnet 3.7, excelling in both coding and reasoning tasks with improved precision and controllability. Achieving state-of-the-art performance across a wide range of applications, Claude Sonnet 4 is designed to handle complex tasks with ease.

by [anthropic](#) | 200K context | \$3/M input tokens | \$1.5M output tokens | \$4.80/K input/imgs

### [Google: Gemini 2.0 Flash](#)

298B tokens

Translation (#1) • Trivia (#1) • Roleplay (#2) • Finance (#2) • Marketing (#2)

Gemini Flash 2.0 offers a significantly faster time to first token (TTFI) compared to Gemini Flash 1.5, while maintaining quality on par with larger models like Gemini Pro 1.5. It introduces notable ...

by [google](#) | 1.05M context | \$0.10/M input tokens | \$0.40/M output tokens | \$0.026/K input/imgs

### [Google: Gemini 2.5 Flash Preview 05-20](#)

205B tokens

Technology (#1) • Marketing/SEO (#1) • Science (#1) • Health (#2) • Trivia (#2) • +7 categories

Gemini 2.5 Flash May 20th Checkpoint is Google's state-of-the-art workhorse model, specifically designed for advanced reasoning, coding, mathematics, and scientific tasks. It includes built-in ...

by [google](#)

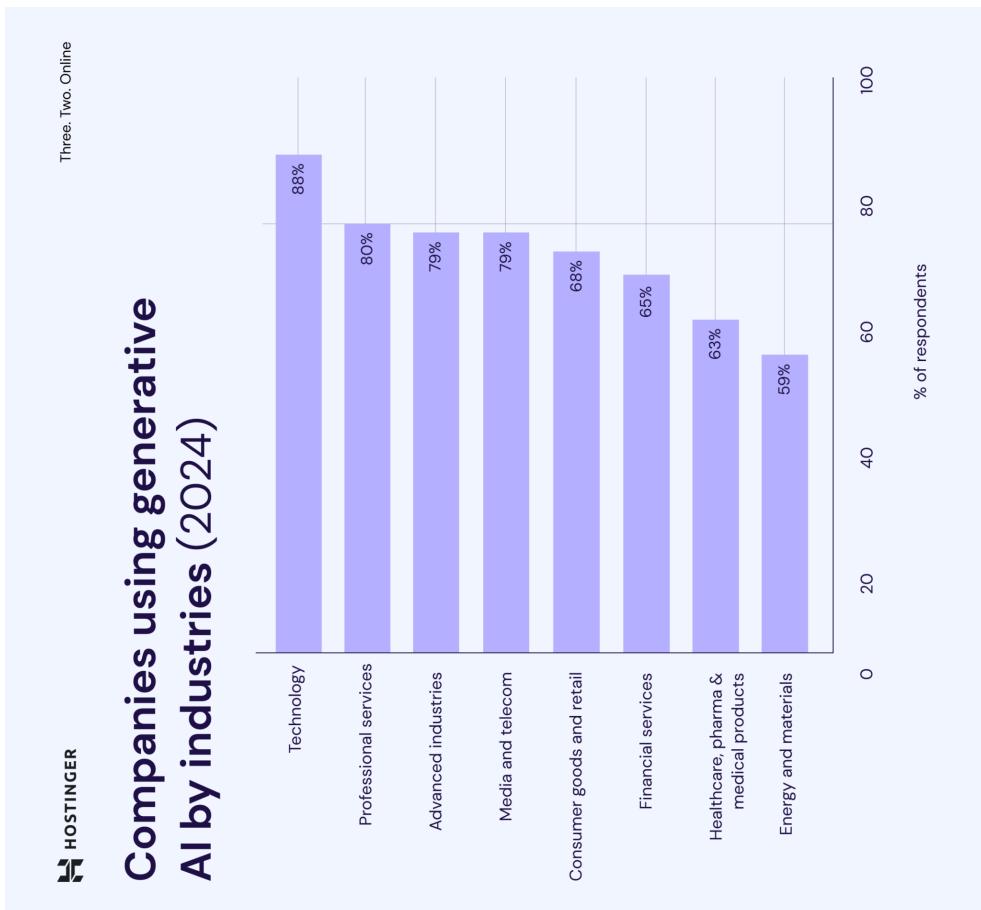
| 1.05M context | \$0.15/M input tokens | \$0.60/M output tokens | \$0.619/K input/imgs

### [DeepSeek: DeepSeek V3 0324 \(free\)](#)

146B tokens

DeepSeek V3, a 685B-parameter, mixture-of-experts model, is the latest iteration of the flagship chat model family from the DeepSeek team. It succeeds the DeepSeek V3 model and performs ...

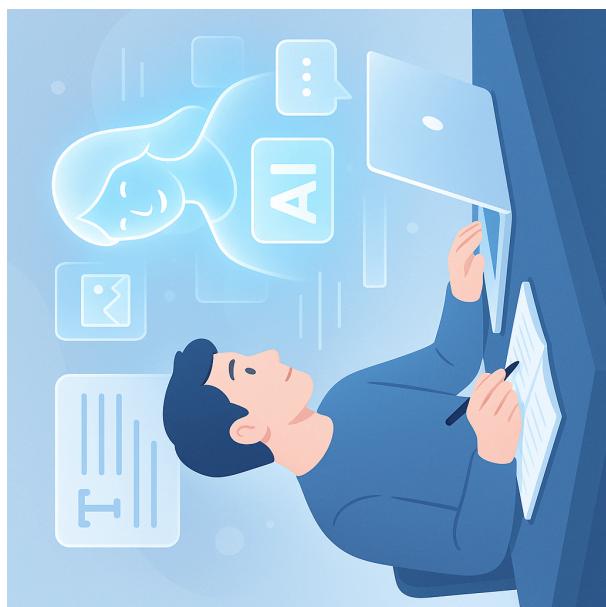
# AI ADOPTION IN DIFFERENT SECTORS



Source: [Hostinger](#)

# THE BENEFICIAL OF LEARNING AI

## 1. Career Resilience & Opportunities





1. Career Resilience & Opportunities
2. Personal Productivity Multiplier



1. Career Resilience & Opportunities
2. Personal Productivity Multiplier
3. Economic Advantage

# **ETHICS & CHALLENGES**

**For Creators/Developers:**

- Design systems to minimize bias in training data
- Implement privacy safeguards and data protection
- Build transparency and explainability features
- Respect intellectual property in training datasets
- Consider environmental impact of model training

**For Users:**

- Verify AI-generated content for accuracy
- Disclose when using AI-generated content
- Avoid using AI-generated content for deceptive purposes (deepfakes, misinformation)
- Respect copyright when prompting with others' work
- Maintain critical thinking rather than blind trust

**Shared Responsibilities:**

- Both must consider impact on jobs and society
- Both should advocate for responsible AI policies
- Both need to understand AI limitations
- Both should prioritize human agency and decision-making

# DEMO

Use Claude Code (Claude CLI) for creating a blog based on this slide

# CONCLUSION

# KEY TAKEAWAYS

- **Generative AI** creates new content (text, images, code) from learned patterns
- LLMs are specialized Gen AI models focused on understanding and generating human-like text
- **Practical Applications** span across industries - from creative automation to customer service
- **Ethical Use** requires both developers and users to prioritize transparency, accuracy, and human agency

# THE FUTURE IS COLLABORATIVE

- AI is a **tool**, not a replacement for human creativity and judgment
- Understanding AI capabilities helps us leverage it effectively in our daily work
  - The key is finding the right balance between automation and human expertise
- Stay curious, keep learning, and adapt responsibly

# THANK YOU!

We Are Hiring!

Scan QR if you are interested joining US



# QNA

