Tribhuvan University
Institute of Engineering
Thapathali Campus, Thapathali

Artificial Intelligence

Submitted by:
Jenish Pant (THA078BEI018)
Sajen Maharjan (THA078BEI048)

Submitted to:
Department of Electronics and Computer Engineering

May 25, 2025

Timeline of Major Breakthroughs and State-of-the-Art (SOTA) Models in AI

Introduction

Artificial Intelligence (AI) has undergone rapid advancements over the past decade, revolutionizing industries and reshaping technological paradigms. This document provides a structured timeline of key breakthroughs and state-of-the-art models across various AI subfields, including computer vision, natural language processing, audio processing, recommender systems, and robotics. By examining these milestones, we gain insights into the evolution of AI and its future trajectory.

1. Computer Vision

Image Classification

- **AlexNet (2012):** Revolutionized deep learning by winning the ImageNet competition with ReLU activations and GPU acceleration.
- **VGGNet (2014):** Demonstrated the effectiveness of deep CNNs with small (3x3) filters.
- **ResNet (2015):** Introduced residual connections, enabling training of networks with 100+ layers.
- **EfficientNet (2020):** Achieved state-of-the-art accuracy via compound scaling of depth, width, and resolution.
- **ConvNeXt (2022):** Modernized CNNs with transformer-inspired designs, matching vision transformer performance.

Image Segmentation

- **Fully Convolutional Networks (2015):** Pioneered end-to-end segmentation by replacing dense layers with convolutions.
- **Mask R-CNN (2017):** Extended Faster R-CNN for instance segmentation with mask prediction.
- **SegFormer (2021):** Lightweight transformer model for efficient segmentation.
- **Mask2Former (2022):** Unified architecture for panoptic/instance/semantic segmentation using masked attention.

Object Detection

- **R-CNN (2014):** Introduced region-based CNN object detection.
- **Faster R-CNN (2015):** Improved efficiency with Region Proposal Networks (RPNs).
- **YOLOv1 (2016):** Framed detection as a regression problem for real-time performance.
- **YOLOv8 (2023):** Current state-of-the-art in speed/accuracy trade-offs.

Image Generation

- **GANs (2014):** Introduced adversarial training for image synthesis.
- **BigGAN (2018):** Scaled GANs to high-resolution, class-conditional generation.
- **DALL·E (2021):** Combined transformers with generative models for text-to-image synthesis.
- **Stable Diffusion (2022):** Open-source diffusion model for controllable generation.
- **DALL·E 3 (2023) / Midjourney v5 (2023):** Advanced text-to-image models with improved prompt adherence.

2. Natural Language Processing (NLP)

Sentiment Classification

- **LSTM (2015):** Improved sequential data modeling for sentiment analysis.
- **BERT (2018):** Introduced bidirectional transformer pretraining.
- **DeBERTa (2021):** Enhanced BERT with disentangled attention mechanisms.

Text Generation

- **GPT-2 (2018):** Demonstrated large-scale transformer capabilities.
- **GPT-3 (2020):** 175B parameter model with few-shot learning.
- **GPT-4 (2023):** Multimodal model with improved reasoning.
- **Claude 3 (2024), Gemini 1.5 (2024), LLaMA 3 (2024):** Next-generation LLMs focusing on safety and multilingual support.

3. Audio

Speech Recognition

- **DeepSpeech (2016):** End-to-end speech-to-text model.
- **Wav2Vec 2.0 (2020):** Self-supervised learning for robust ASR.
- **Whisper (2022):** Multilingual model robust to noise and accents.

Text-to-Speech

- **Tacotron 2 (2017):** Neural TTS with WaveNet vocoder.
- **VALL-E (2022):** Zero-shot voice cloning.
- **Bark (2023):** Expressive, controllable TTS.

4. Recommender Systems

- **Matrix Factorization (2006):** Classic collaborative filtering.
- **Wide & Deep (2016):** Hybrid memorization/generalization model.
- **LLM-based RecSys (2023):** Leveraged language models for personalized recommendations.

5. Robotics

- **OpenAI Gym (2016):** Standardized RL benchmarks.
- **Boston Dynamics Spot (2020):** Agile quadruped robot.
- **RT-2 (2023):** Vision-language-action model for robotic control.

Conclusion

The AI landscape continues to evolve rapidly, with each breakthrough building on prior work. From convolutional networks to multimodal LLMs, these advancements highlight the field's trajectory toward human-like capabilities. Future progress will likely focus on efficiency, safety, and real-world integration.