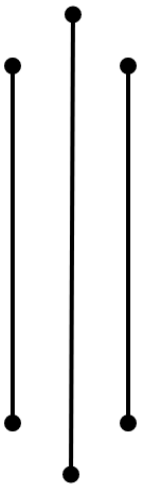
Tribhuvan University
Institute of Engineering
Thapathali Campus, Thapathali

Artificial Intelligence



Submitted by:
Dishan Shakya (THA078BEI014)
Diwas Dahal (THA078BEI015)

Submitted to:
Department of Electronics and Computer Engineering

May 25, 2025

Timeline of Key Innovations and Cutting-Edge AI Models

Introduction

Artificial Intelligence (AI) has seen unprecedented growth in recent years, reshaping how technology is applied across numerous fields. This report outlines a chronological overview of pivotal developments and state-of-the-art models across several AI domains such as computer vision, natural language processing, audio, recommendation systems, and robotics. The timeline not only illustrates progress but also hints at the direction AI is heading.

1. Computer Vision

Image Classification

- **AlexNet (2012):** Marked a significant leap in deep learning, leveraging GPUs and ReLU for performance.
- **VGGNet (2014):** Validated the use of deeper CNNs with uniform, small-sized filters.
- **ResNet (2015):** Solved training degradation via residual connections, enabling deeper architectures.
- **EfficientNet (2020):** Introduced compound scaling to balance network depth, width, and resolution.
- **ConvNeXt (2022):** Brought transformer-like innovations into CNNs, narrowing the performance gap.

Image Segmentation

- **Fully Convolutional Networks (2015):** Enabled full-image segmentation by using only convolutional layers.
- **Mask R-CNN (2017):** Added pixel-wise mask prediction to Faster R-CNN for instance-level segmentation.
- **SegFormer (2021):** Provided an efficient transformer-based approach for segmentation.
- **Mask2Former (2022):** Unified different segmentation tasks using masked attention mechanisms.

Object Detection

- **R-CNN (2014):** Pioneered region-based object detection using CNNs.
- **Faster R-CNN (2015):** Boosted detection speed by integrating Region Proposal Networks.
- **YOLOv1 (2016):** Reframed detection as a regression task for real-time applications.
- **YOLOv8 (2023):** Represents the current benchmark in balancing accuracy and speed.

Image Generation

- **GANs (2014):** Introduced adversarial training for generating realistic images.
- **BigGAN (2018):** Scaled GANs for higher-resolution image generation.
- **DALL·E (2021):** Combined transformers and generation for text-based image synthesis.
- **Stable Diffusion (2022):** Open-source model offering controllable image generation.
- **DALL·E 3 / Midjourney v5 (2023):** Enhanced prompt fidelity and artistic quality in generated visuals.

2. Natural Language Processing (NLP)

Sentiment Analysis

- **LSTM (2015):** Improved contextual understanding in sentiment classification.
- **BERT (2018):** Enabled deep bidirectional context with masked language modeling.
- **DeBERTa (2021):** Refined BERT's architecture using disentangled attention for better performance.

Text Generation

- **GPT-2 (2018):** Demonstrated language modeling at scale.
- **GPT-3 (2020):** Set a new bar with 175B parameters and few-shot capabilities.
- **GPT-4 (2023):** Introduced multimodal reasoning and greater reliability.
- **Claude 3, Gemini 1.5, LLaMA 3 (2024):** Represent a new generation of LLMs with improvements in alignment and multilingual abilities.

3. Audio Processing

Speech Recognition

- **DeepSpeech (2016):** End-to-end architecture for automatic speech recognition.
- **Wav2Vec 2.0 (2020):** Self-supervised model that improves robustness and efficiency.
- **Whisper (2022):** Handles multiple languages and performs well under noisy conditions.

Text-to-Speech

- **Tacotron 2 (2017):** Neural model that pairs with WaveNet for lifelike speech.
- **VALL-E (2022):** Capable of zero-shot voice cloning from limited samples.
- **Bark (2023):** Offers nuanced and expressive synthetic speech output.

4. Recommender Systems

- **Matrix Factorization (2006):** Laid the foundation for collaborative filtering techniques.
- **Wide & Deep (2016):** Combined memorization with generalization for better recommendations.
- **LLM-driven Recommenders (2023):** Leveraged large language models to personalize user suggestions.

5. Robotics

- **OpenAI Gym (2016):** Provided a common toolkit for reinforcement learning experimentation.
- **Boston Dynamics Spot (2020):** Demonstrated real-world agility in quadruped robots.
- **RT-2 (2023):** Integrated vision and language models to control robotic actions effectively.

Conclusion

The evolution of AI is marked by continuous innovation, each advancement paving the way for more sophisticated systems. From early CNNs to today's versatile LLMs, the progress underscores AI's growing capacity to handle complex, real-world problems. Looking forward, the focus is shifting toward safer, more efficient, and practically deployable solutions.