Introduction to Generative AI

Fariba Yousefi

April 7, 2025

Outline

Brief History of AI [1]

Early AI (1950s–1980s)

- 1950: Turing proposes the Turing Test.
- 1955: Dartmouth AI Conference – Birth of AI.
- 1959: First self-learning checkers-playing AI; “machine learning” coined.
- 1980s: AI “Winter” due to symbolic AI limitations.

[1] IBM History of AI

Neural Network Resurgence (1990s–2000s)

- 1986: Backpropagation revitalizes deep networks.
- 1997: LSTMs improve sequence learning.
- 2012: AlexNet wins ImageNet, showcasing deep learning.

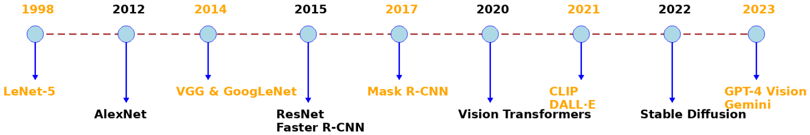
Generative AI Emerges (2010s)

- 2013: Introduction of Variational Autoencoders (VAEs).
- 2014: GANs (Generative Adversarial Networks) are introduced.
- 2017: Transformers replace RNNs, advancing language models.
- 2018: BERT & GPT models transform NLP.

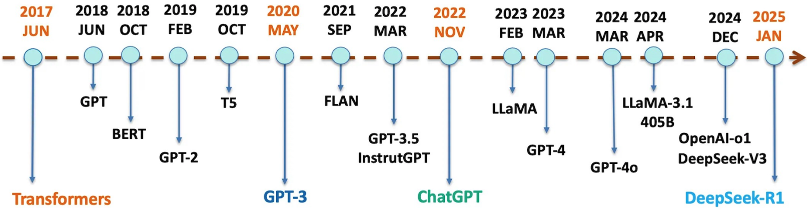
AI Scaling Up (2020s–Present)

- 2020: GPT-3 launches demonstrating large-scale language modeling.
- 2021: DALL-E generates images from text.
- 2022: Stable Diffusion democratizes AI art.
- 2023: GPT-4, Bard, and multimodal AI push boundaries.

A Brief History of Vision models



A Brief History of LLMs



What Are Foundation Models?

- **Definition:** Large-scale AI models pre-trained on massive datasets.
- General-purpose models that can be fine-tuned for various tasks.

Key Characteristics of Foundation Models

- Pretrained on massive, Internet-scale datasets.
- General-purpose and adaptable via fine-tuning.
- Highly scalable with billions of parameters.
- Employ transfer learning and support multimodal input.

Discussion

How do foundation models differ from traditional AI models?

Discussion: How do foundation models differ from traditional AI models?

- Generalization vs. Specialization.
- Scale of Training.
- Self-Supervised Learning.
- Transferability.

Examples of Foundation Models

Model	Type	Key Application
GPT-4 (OpenAI)	Language Model	Text generation, summarization, chatbots
BERT (Google)	Language Understanding	Search engines, text classification
CLIP (OpenAI)	Vision-Language Model	Image classification, text-image retrieval
DALL·E (OpenAI)	Generative Vision Model	Image generation from text prompts
Stable Diffusion	Diffusion Model	AI-generated art, scientific visualization
AlphaFold (DeepMind)	Scientific Model	Protein structure prediction

Discussion

What is the key difference between BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer)?

- BERT: Processes text bidirectionally, meaning it looks at both left and right contexts when understanding language.
- GPT: Uses an autoregressive (left-to-right) approach, predicting the next word based on previous words, making it excellent for text generation tasks.

How Foundation Models Work

- **Pretraining Phase:** Self-supervised learning from large datasets.
- **Fine-Tuning Phase:** Customizing models for specific tasks.
- **RLHF:** Reinforcement Learning from Human Feedback for alignment.

Discussion

Why does self-supervised learning work well for foundation models?

Discussion: Why does self-supervised learning work well for foundation models?

- Removes the Need for large-labeled Datasets.
- Enables Generalization.
- Scalability.
- Rich Representations.

Applications of Foundation Models

- **NLP:** Chatbots, summarization, translation.
- **Computer Vision:** Image generation, medical imaging.
- **Code Generation:** AI-assisted programming (e.g., GitHub Copilot).
- **Drug Discovery:** Protein structure predictions.
- **Multimodal AI:** Combining text, images, and audio.

Discussion

Why is fine-tuning crucial for domain-specific AI applications?

Discussion: Why is fine-tuning crucial for domain-specific AI applications?

- Improves Domain-Specific Accuracy.
- Reduces Errors & Hallucinations.
- Enhances Performance on Low-Resource Domains.
- Ensures Compliance with Industry Standards.
- Reduces Computational Costs Compared to Training from Scratch.

Self-Supervised Learning (SSL)

What is it? Learning from unlabeled data by predicting parts of the input using other parts. No human-annotated labels required.

Key Ideas

- Masked token/image prediction - mask part of the input and train the model to reconstruct it.
  Example: Masked Auto Encoders (MAE), BERT.
- Contrastive learning - Learn representations by comparing positive (similar) and negative (dissimilar) pairs.
  Example: SimCLR.
- Predictive pretext tasks - Design artificial tasks like predicting rotation, patch order.
  Example: RotNet (predict image rotation angle).

Why SSL Matters

Why It Matters

- Uses massive unlabeled datasets
- Scales to real-world problems
- Enables general-purpose foundation models

Paradigm	Labels required	Main Objective
Supervised	Yes	Learn from labeled data
Unsupervised	No	Find structure/pattern in data
Self-Supervised	No	Learn from data
Reinforcement	Depends	Learn from feedback signals (reward)

What is deep Generative AI?

- Generative AI creates new data similar to its training data
- It learns the underlying patterns in data
- Can generate text, images, audio, etc.
- Example: ChatGPT generating text responses

Generative vs. Discriminative Models

- Discriminative: Classify or label data (e.g., cat vs. dog)
- Generative: Create new samples from the data distribution
- Example:
  - Discriminative: Is this a cat?
  - Generative: Generate a new image of a cat

Discussion

Can you think of examples where a generative approach would be more useful than a discriminative one, and vice versa?

What is something creative or useful you would want a generative AI to do?

Core Concepts of Generative Modeling

- Probability Distributions & Sampling
- Latent Space
- Training Paradigms

Probability Distributions

- Generative models learn the probability distribution of data
- They sample new data points from this distribution
- **Analogy:** A jar of multicolored beads — sampling mimics observed ratios

Latent Space

- A compressed, abstract representation of data
- Captures essential features (e.g., smile intensity, age, color tone)
- Smooth transitions in latent space create smooth transitions in output
- **Analogy:** A library catalog with key themes instead of full books

Training Generative Models

- Learn from large datasets without explicit labels
- Models try to generate realistic outputs and are rewarded when successful
- Loss functions measure the error or difference between generated data and real data

Discussion

What is the difference between foundation models and generative AI?

Variational Autoencoders (VAEs)

- Encode data into a latent space, then decode it back
- Generates new data by sampling from latent space
- **Strength:** Stable to train
- **Limitation:** Output lack sharpness or fine detail
- **Applications:** Data compression, anomaly detection, generation

Discussion

VAEs intentionally add randomness during training (in the latent space encoding). Why do you think injecting some noise (randomness) might actually help a generative model create better outputs? What could go wrong if the model simply memorized how to compress and decompress each training example perfectly?

Discussion: What if the model just memorized?

- Noise: Helps the model generalize.
- Noise: It encourages the latent space to be smooth and continuous.
- No noise: Model loses its ability to generalize.
- No noise: The latent space would be disconnected and overfit.

Generative Adversarial Networks (GANs)

- Generator creates fake data; Discriminator learns to distinguish real from fake
- They train in competition
- **Applications:** Realistic image generation, super-resolution

Discussion

GANs pit two AI agents against each other. In what ways might this adversarial setup be advantageous for learning, and what dangers might it pose if, say, the discriminator far outperforms the generator (or vice versa)? Can you draw a parallel to any real-world learning scenario involving competition?

Discussion

- **Disadvantages:**
  - If one side gets too strong, learning collapses.
  - Discriminator: Easily spots fakes, can't improve.
  - Generator: Easily fools discriminator, no useful critique.
  - Result: Unstable training, a common challenge with GANs (e.g., mode collapse, vanishing gradients).

Diffusion Models

- Learn to reverse a noising process step-by-step
- Start from noise, gradually denoise to produce an image
- **Applications:** Text-to-image models (e.g., DALL-E, Stable Diffusion)

Discussion

Diffusion models learn by destroying data and then recovering it. Why might learning to undo noise results in a powerful generation ability? Can you connect this to any intuitive idea of how humans or other systems handle uncertainty or incomplete information?

Discussion

- Learning to undo noise teaches a model to recognize the underlying structure of data — what's essential and what's random.
- The model doesn't just memorize exact samples — it learns the patterns and dependencies that define realistic data.
- By mastering the art of denoising, it gains the ability to reconstruct or imagine new samples that fit the learned structure, even from randomness.

Transformers and LLMs

- Self-attention allows modeling long sequences
- Used in language generation (GPT), code, and image tasks
- **Applications:** Chatbots, summarization, code generation

Discussion

Large language models can generate human-like text. What are some situations where you would not want to trust a language model's generated output without verification? Consider things like accuracy, bias, or responsibility in your answer.

Discussion

- Factual Accuracy: LLMs can generate plausible-sounding but incorrect information (hallucinations).
- Bias and Fairness: LLMs are trained on vast internet data — including biased or harmful content.
- Responsibility & Accountability: If a model gives bad advice or causes harm, who is responsible?

Summary Table

	Variational Autoencoders (VAEs)	Generative Adversarial Networks (GANs)	Diffusion Models	Transformers
Architecture	Encoder-Decoder Structure	Generator and Discriminator	Forward and Reverse Processes	Self-Attention Mechanism
Training Objective	Maximize Evidence Lower Bound (ELBO)	Minimax game	Denoising Score Matching	Sequence Modeling
Loss Function	Evidence Lower Bound (ELBO)	Generator Loss Discriminator Loss	Denoising Score Matching Loss	Cross-Entropy Loss
Explanation of Loss Function	ELBO: -Reconstruction Loss -Regularization Term (KL Divergence)	Generator Loss Discriminator Loss	Denoising Score Matching Loss	Cross-Entropy Loss
Strengths	- Stable Training - Latent Space Representation - Explicit Likelihood Estimation	- High-Quality Outputs - Flexible Architecture	- High Fidelity and Diversity - Stable Training Dynamics	- Handles Long-Range Dependencies - Scalability
Limitations	- Blurry Outputs - Limited to Simple Data	- Training Instability - Mode collapse	- Computationally Intensive - Slow Sampling	- Resource Demanding - Data Hungry
Typical Applications	- Anomaly Detection - Data Compression - Feature Learning	- Image Synthesis - Data Augmentation - Super-Resolution	- Text-to-Image Generation - Image and Video Synthesis	- Machine Translation - Text Summarization - Text Generation - Question Answering

Applications of Generative AI

- Art and Design (image, video, 3D)
- Writing, summarization, and chatbots
- Music and speech generation
- Drug discovery and scientific research

Creative Industries

- Text-to-image for concept art and prototyping
- AI music assistants for composition and remix
- Style transfer and video editing

Science and Healthcare

- AI-designed molecules and proteins
- Synthetic data for medical research
- Generative design in engineering

Bias and Fairness

- AI can replicate and amplify societal biases
- Important to audit data and outputs
- Equity in model deployment matters

Hallucinations and Deepfakes

- Models may generate plausible but incorrect info
- Deepfakes pose misinformation threats
- Critical thinking and verification are essential

Discussion

What are the challenges in integrating AI-generated research findings?

Discussion: What are the challenges in integrating AI-generated research findings?

- Verification and Accuracy.
- Lack of Explainability.
- Regulatory and Ethical Considerations.
- Computational Costs.

Intellectual Property and Privacy

- Who owns AI-generated content?
- Training on copyrighted or private data raises concerns
- Legal and policy frameworks are evolving

Jobs and Energy Use

- Creative and knowledge work may be disrupted
- Need for upskilling and responsible deployment
- Training large models requires significant energy

What's Next in Generative AI?

- Multi-modal models (text + image + audio)
- Real-time content generation
- Personalized and domain-specific models
- Regulatory and safety innovations

Discussion

What most excites or concerns you about using generative AI in scientific research?

Discussion

How will you use generative AI responsibly and creatively in your research or work?

Conclusion

- Generative AI is powerful and transformative
- Key techniques: VAEs, GANs, Diffusion, Transformers
- Wide applications, real challenges
- Ethical and critical use is essential