Large Language Models

The Foundation of Modern NLP

ITAI2373 – Module 10

From N-grams to Transformers

Learning Outcomes

Understand probabilistic foundations of language modeling

Compare N-gram models with neural approaches

Explore transformer architecture and pre-trained models

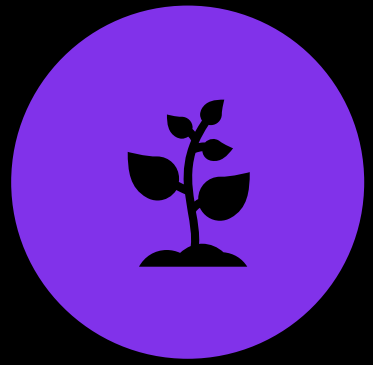
Recognize ethical implications and biases

Our Journey Today

- Foundations: What are Language Models?
- N-gram Models: The Classic Approach
- Neural Revolution: RNNs to Transformers
- Modern Giants: BERT, GPT, and Beyond
- Ethics and Responsibility

Building on Our Foundation

TEXT PREPROCESSING ENABLES MODEL TRAINING



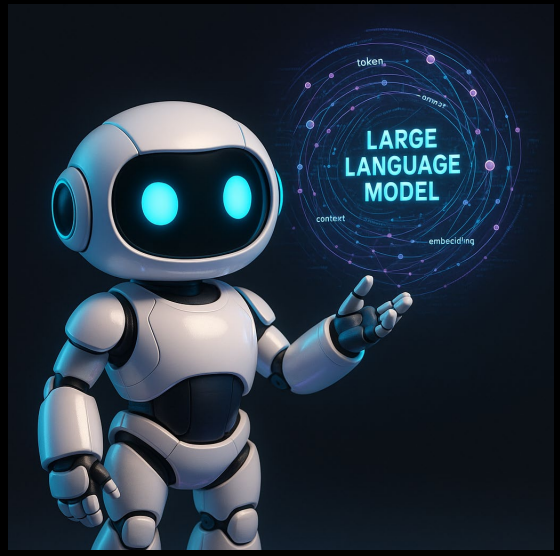
WORD EMBEDDINGS FROM MODULE 4 ARE CRUCIAL

CLASSIFICATION AND NER USE LANGUAGE MODEL FEATURES

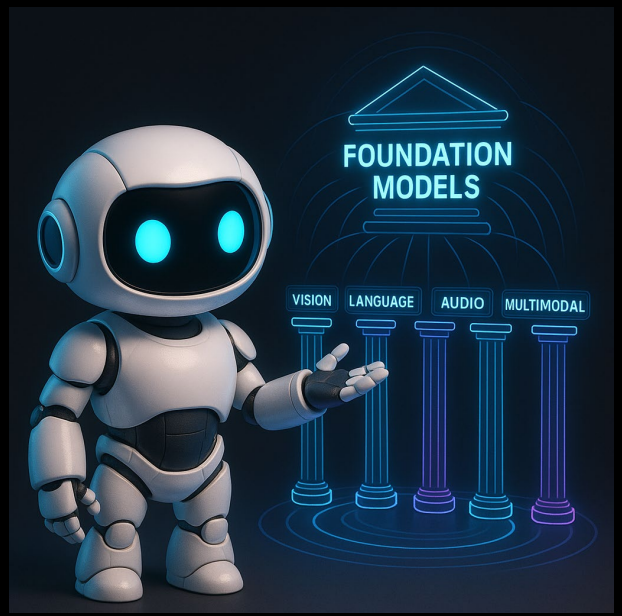
What is a Language Model?

LLMs are deep-learning-based models that use many parameters to learn from vast amounts of unlabeled texts. They can perform various natural language processing tasks such as recognizing, summarizing, translating, predicting, and generating text.

- System that learns the probability of word sequences
- Predicts the next word given previous context
- Foundation of most NLP applications



Foundation Models vs LLMs



Traditional ML

Training *Tasks*

AI1
AI2
AI3
AI4
AI5
AI6

- Individual siloed models
- Require task-specific training
- Lots of human supervised training

Foundation Models

Massive external data

Massive Foundation Model

Prompting

Q&A
Translation
...

Enterprise proprietary data

Fine Tuned Model

Prompting

Classification
Code Gen
...

- Massive multi-tasking model
- Adaptable with little or no training
- Pre-trained unsupervised learning

LLM: The Prediction Game

- Given: 'The weather today is...'
- Possible completions: sunny, rainy, cold, beautiful
- Model assigns probabilities to each option

Language Models Power Everything

AUTOCOMPLETE AND TEXT SUGGESTIONS

MACHINE TRANSLATION

CHATBOTS AND VIRTUAL ASSISTANTS

CONTENT GENERATION AND SUMMARIZATION

N-gram Magic in Action

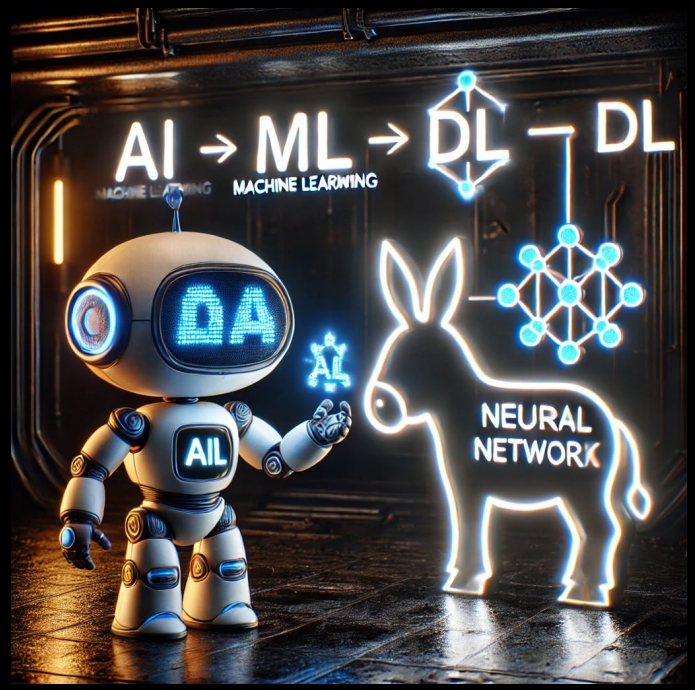
- Foundation of classical NLP
- Training: Count word sequ
  large text
- Prediction: Find most frec
  continuation
- Example: 'I love' → 'you' (m
  completion)

Where N-grams Struggle

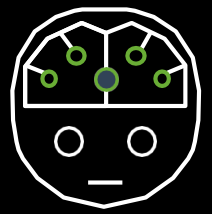
- Limited context window
- Sparse data problems
- No understanding of meaning

Where Do Neural Networks Fit in AI/ML?

- **AI:** Making machines perform human-like intelligence tasks
- **ML:** Enabling machines to learn from data
- **Neural Networks:**
  - Fundamental building blocks of deep learning
  - Capable of learning complex patterns from data
  - Used in Computer Vision (CV), NLP, and autonomous systems



AI, ML, Deep learning?



Artificial intelligence (AI)

Any technique that allows computers to mimic human intelligence using logic, if-then statements, and machine learning

Machine learning (ML)

A subset of AI that uses machines to search for patterns in data to build logic models automatically

Deep learning (DL)

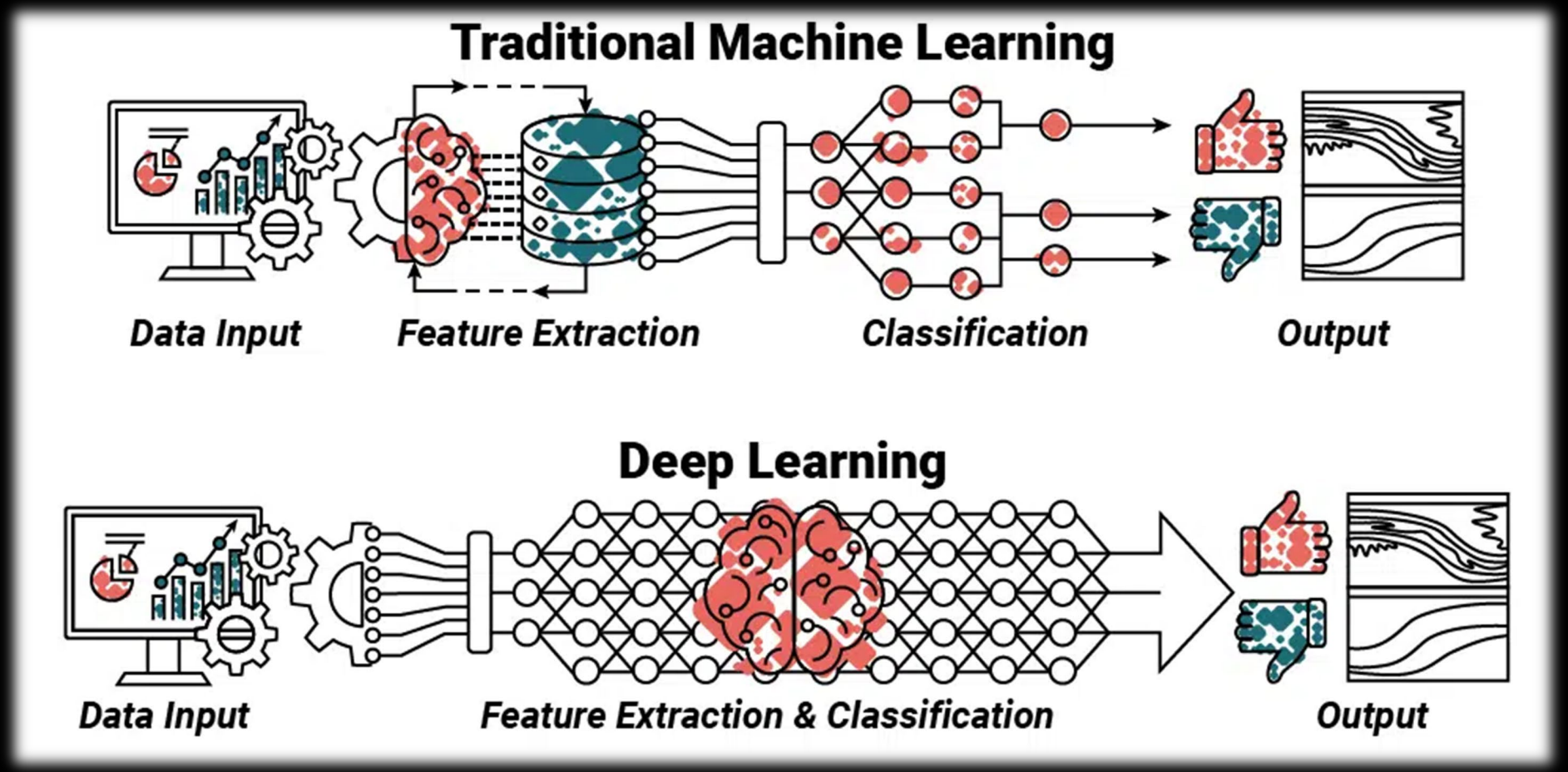
A subset of ML composed of deeply multi-layered neural networks that perform tasks like speech and image recognition

Generative AI

Powered by large models that are pretrained on vast corpuses of data and commonly referred to as foundation models (FMs)

aws

Traditional ML vs Deep Learning

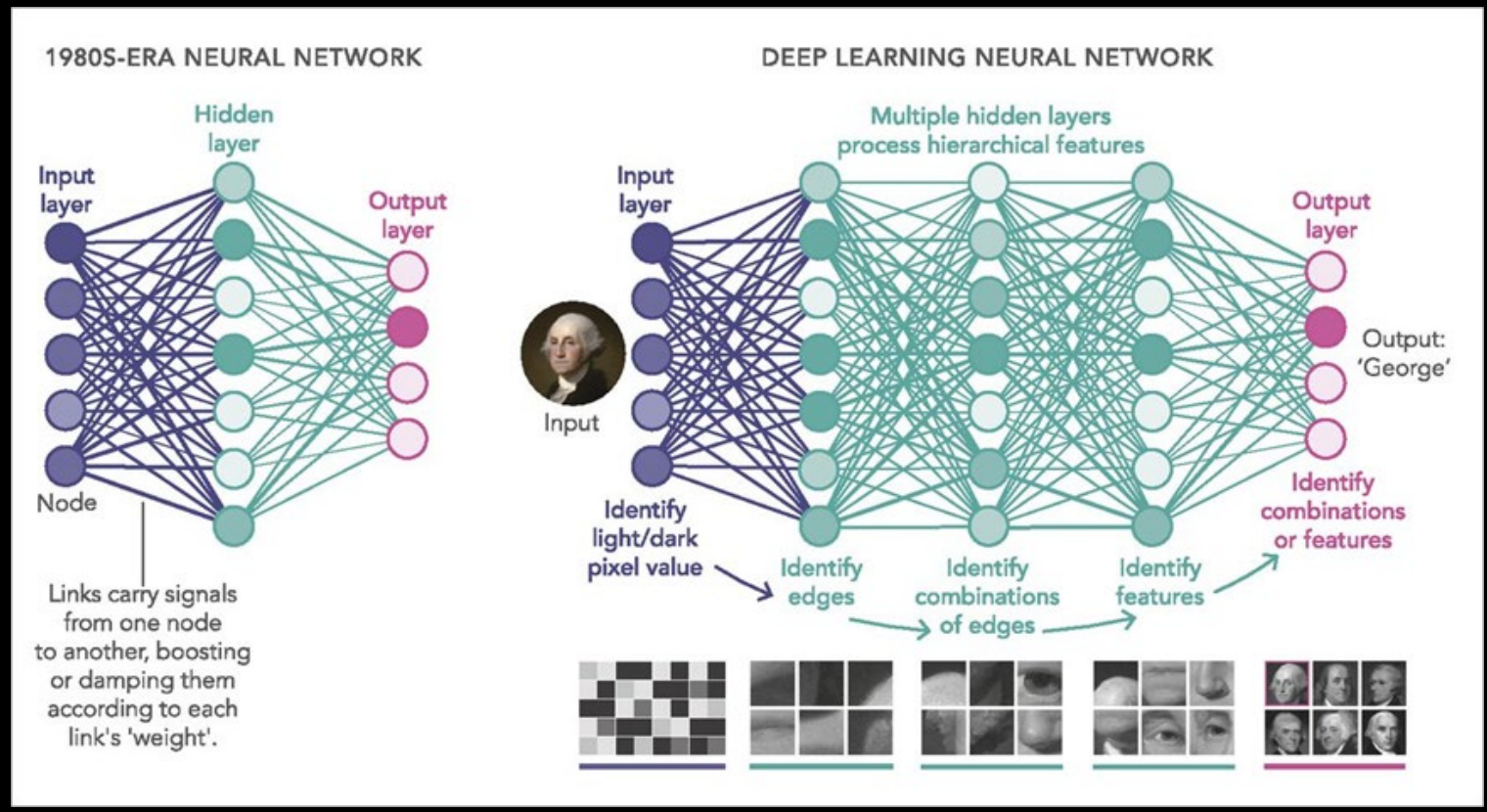


ML vs. LLM (Deep Learning)

- 🔢 **Data Size & Type**
  - **LLM**: Built for massive, unstructured datasets (text, code, documents).
  - **ML**: Suited for smaller, structured data (e.g., tables, CSVs).
- 🧠 **Task Complexity**
  - **LLM:** Handles complex, language-based tasks—generation, reasoning, summarization.
  - **ML:** Best for well-defined problems—classification, regression, predictions.
- ⚙️ **Infrastructure Needs**
  - **LLM:** Requires high compute (GPUs, TPUs, cloud APIs).
  - **ML:** Runs efficiently on standard or local machines.
- 🔍 **Interpretability**
  - **ML**: Easier to explain (e.g., decision trees, linear models).
  - **LLM:** Powerful but opaque—often a black box.

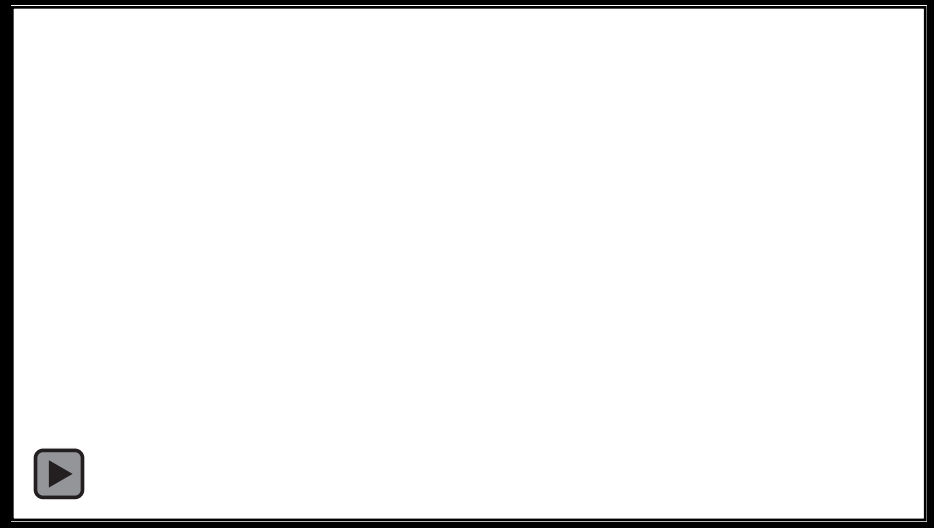
Enter Neural Networks

- Learn dense word representations
- Capture longer-range dependencies
- Understand semantic relationships



Neural Networks Key components

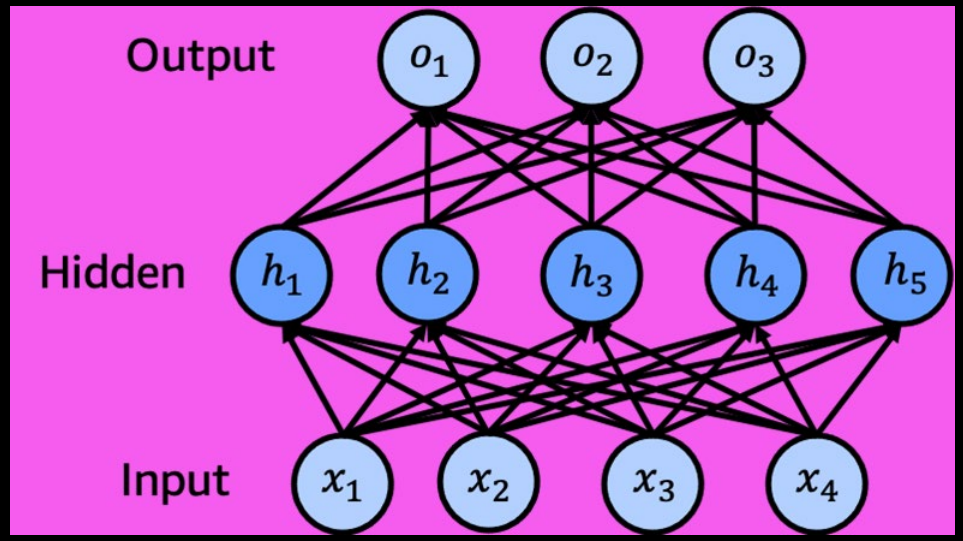
- Artificial Neuron(Perceptron)
  - Activation Functions
  - Weights & Bias
- Layers (Architecture)
  - Input layer
  - Hidden layers
  - Output layer
- Backpropagation
- Gradient Descent
- Cost Function



Feedforward Neural Networks:

- Input data flows in one direction, from input layer to output layer
- Each input is treated independently
- **No memory of past inputs**
- Great for tasks dealing with fixed-size inputs and isolated data points

Feed-forward: Information travels from one layer to the next.



Sequential Processing: RNNs & LSTMs

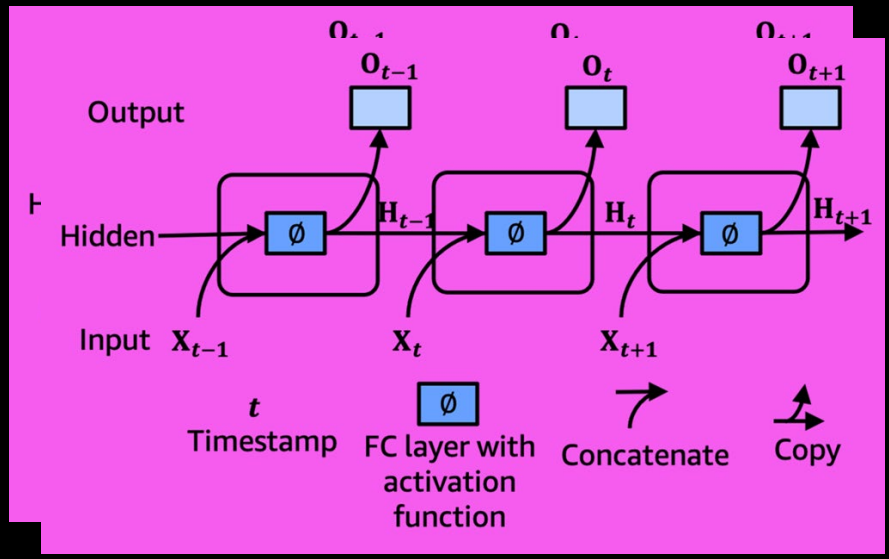
RNNs

- Process text one word at a time
- Remember previous words
- Handle long-range dependencies well

LSTMs

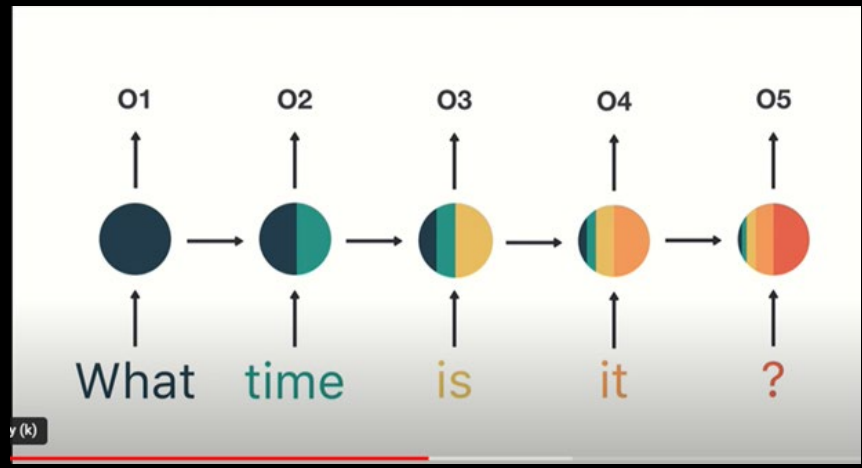
- Addresses RNN Limitation

Recurrent:
Information can travel from one unit to another within the same layer.



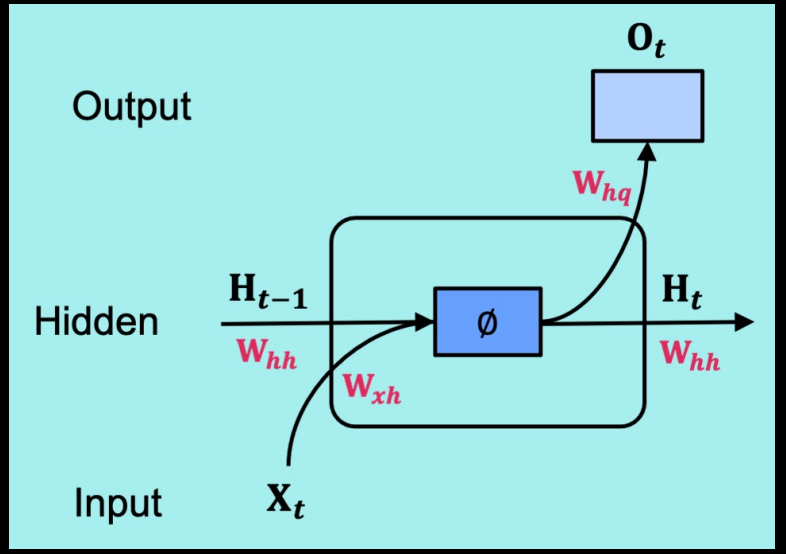
RNN’s & LSTM’s Limitations

- **Vanishing Gradients**: RNNs struggle to retain long-term dependencies, making it hard to learn from earlier inputs.
- **Sequential Bottleneck**: RNNs process toke s one at a time, limiting speed and parallelism.
- **LSTMs**: Improve memory but still rely on sequential flow, constraining scalability and long-range learning.



Simple Recurrent Unit (SRU)

- **SRU** is a simplified version of an RNN, capturing the core concept of using memory for sequential tasks.
- It's like having a short-term memory that helps understand the flow of information.
- SRUs use mathematical functions to perform these actions.



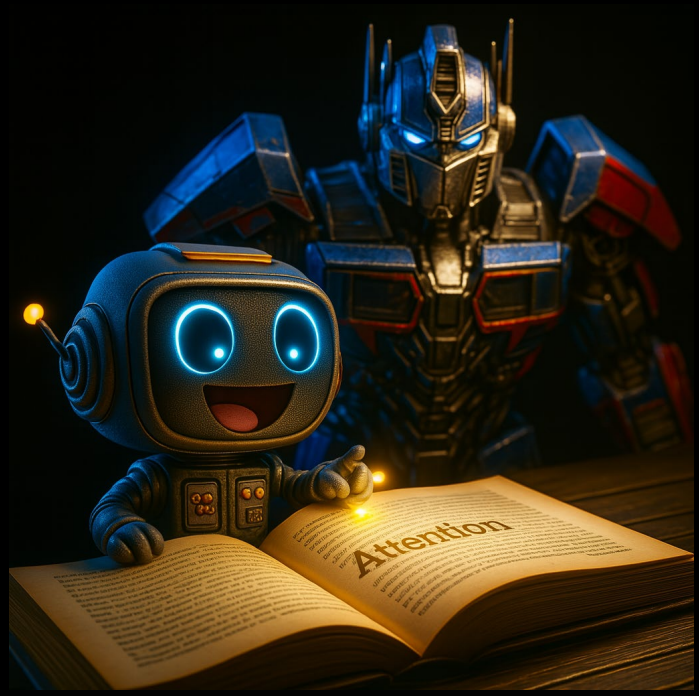
Transformers: The Game Changer

Attention: The Secret Sauce

- Process all words simultaneously
- Attention mechanism focuses on relevant parts
- Foundation of modern NLP

Attention Mechanism

- Focuses on relevant words for each prediction
- Learns what to pay attention to
- Enables understanding of complex relationships



BERT: Bidirectional Understanding

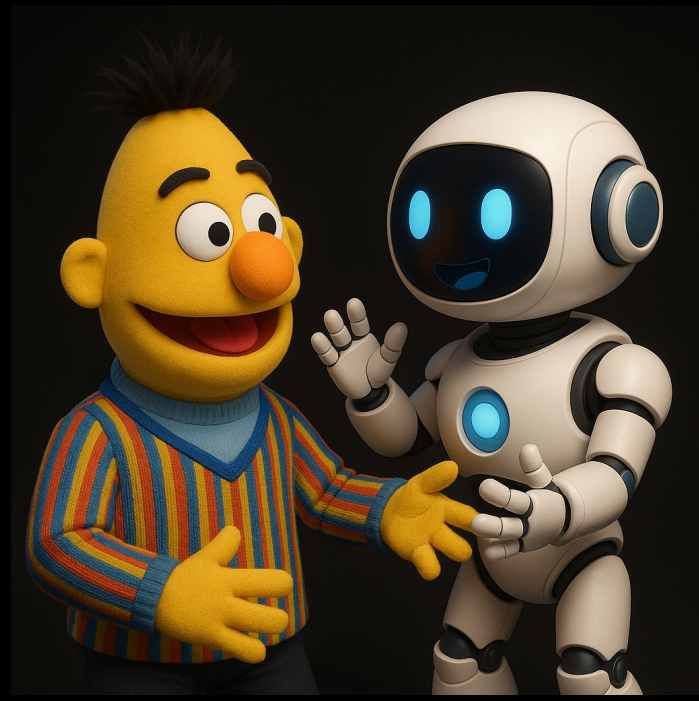
↔ Bidirectional context from both directions

🎭 Masked Language Modeling training objective

🔗 Next Sentence Prediction for coherence

🎯 Pre-training + Fine-tuning paradigm

🏆 State-of-the-art on 11 NLP tasks at release

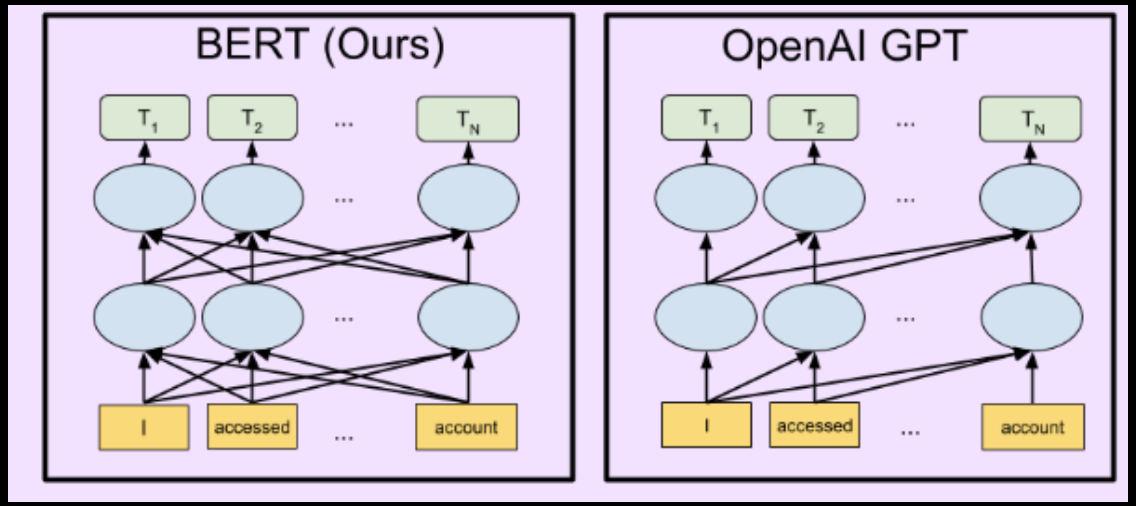


GPT: The Generation Master (2022)

- Autoregressive text generation
- Excels at creative and conversational tasks
- Foundation of ChatGPT and similar systems

Different Strengths

- BERT: Understanding and analysis
- GPT: Generation and conversation
- Choose based on your task needs



The Pre-training Revolution

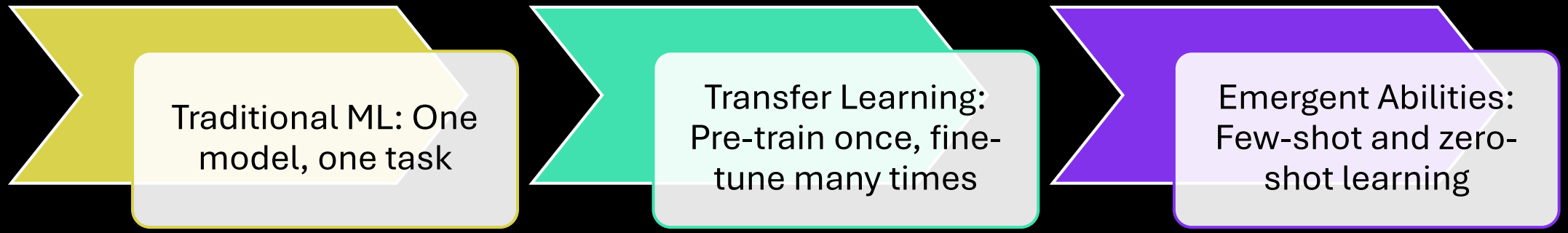
- Train once on massive text
- Fine-tune for specific tasks
- Democratizes access to powerful models

Evolution of Model Learning

Understanding Model Capabilities

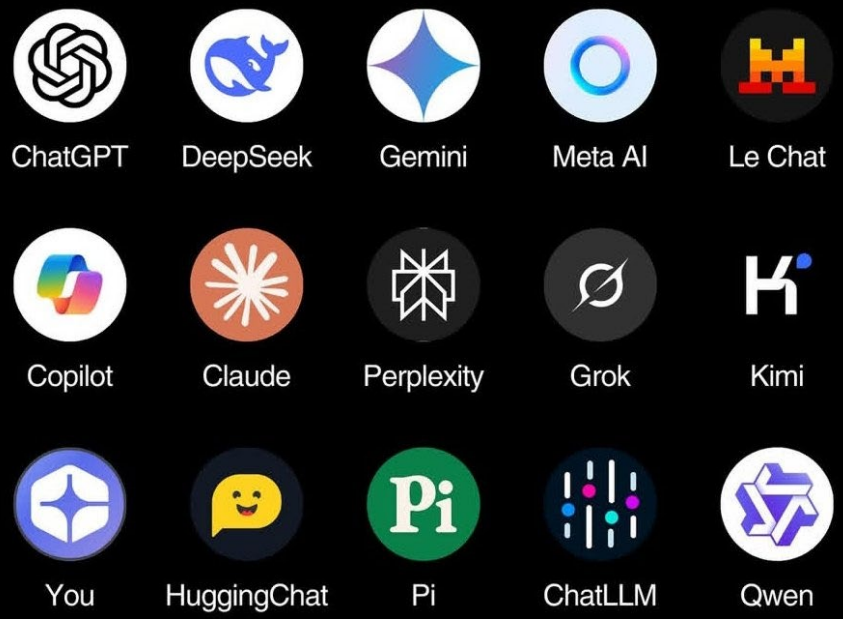
Three levels of Model application

- Traditional: Task-specific training required
- Transfer Learning: Fine-tuning on small datasets
- Few-shot Learning: Minimal task-specific examples needed



LLM’s Today

- GPT-x, Claude, Gemini, Grok
- Billions of parameters
- Multimodal capabilities
- **Strengths:** Language understanding, generation, reasoning
- **Limitations:** Hallucinations, knowledge cutoffs, computational cost



With Great Power...

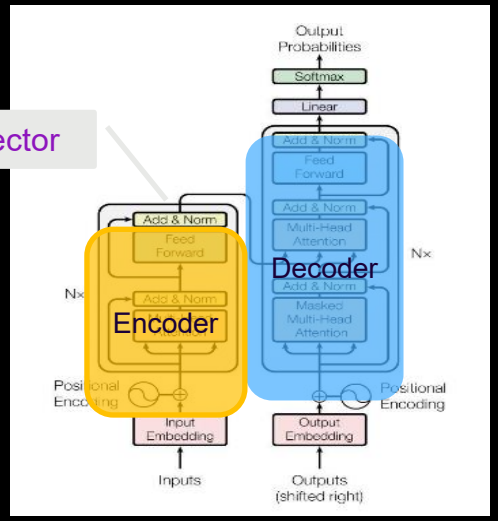
- **Bias and fairness issues**
  - Models learn from human-generated text
  - Inherit societal biases and stereotypes
  - Can amplify existing inequalities
- **Misinformation and deepfakes**
- **Privacy and data concerns**
- **Guardrails:**
  - Diverse training data and teams
  - Bias detection and mitigation
  - Transparency and accountability

LLM and NLP The Link - 2019

- Step 1: Pre-training a Transformer
- Step 2. Fine tune for a specific task

Output 1: Reconstruct missing words
family, of, this, the, Louis, personally, telephone

High dimensional vector



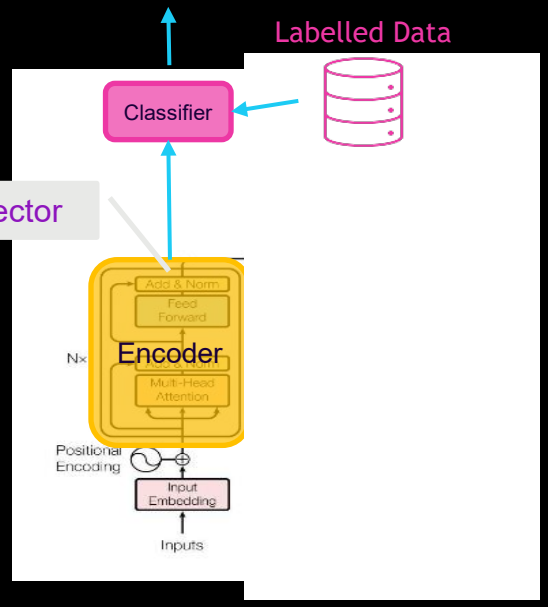
Input: Two sentences with 15% of words masked out

1 = "Initially he supported himself and his family by farming on a plot of family land."

2 = "This in turn attracted the attention of the *St. Louis Post-Dispatch*, which sent a reporter to Murray to personally review Stubblefield's wireless telephone."

Labelled Data

High dimensional vector



LLM's Today & Future Direction

APIs and cloud services

Fine-tuning for specific tasks

Integration with existing systems

Larger, more efficient models

Better reasoning and planning

Multimodal integration

KEY TAKEAWAYS

Language models predict word sequences probabilistically

Evolution from N-grams to transformers represents major progress

Modern models like BERT and GPT have different strengths

Ethical considerations are crucial for responsible development