

# Deep Transfer Learning

## Philosophy of ML & AI in Literary Traditions

Ashley Gao

William & Mary

March 05, 2025

# Overview

- In this lecture we will explore some of the philosophical aspects of artificial intelligence and machine learning **in literary traditions**.
  - Which means we will not cover Nietzsche, Kant, Sartre ...
  - But we will (briefly) explore how people viewed the concept of artificial life in different historical periods and in different cultures.
- Think of this lecture as a thought experiment!
- (And obviously you will not be tested on these.)

# AI & ML in Ancient Times

- The idea of AI & ML can be traced to ancient Greece & Rome (but it is not clear where it originated)!



- Hephaestus, who created Talos



- Pygmalion, who created Galatea

# Hephaestus and Talos

- ... *not in the whole of the kingly art, but in one subsidiary to the kingly, enough for presiding in law courts; so that he was spoken of as a good judge. For Minos used him as guardian of the law in the city, and Talos as the same for the rest of Crete. For Talos thrice a year made a round of the villages, guarding the laws in them, by holding their laws inscribed on brazen tablets, which gave him his name of “brazen.” - Plato, Minos*

# Hephaestus and Talos

- Hephaestus, who created Talos
  - Creator and Creation
  - Autonomy and Free Will
  - The Role of Technology in Society
- Talos, created by Hephaestus
  - Artificial Sentience
  - Human-Machine Boundaries
  - Morality and Ethics
  - The Problem of Control

# Pygmalion and Galatea

- ... *Pygmalion had seen them (the Propoetides), spending their lives in wickedness, and, offended by the failings that nature gave their hearts, he lived as a bachelor, without a wife or partner for his bed. But, with wonderful skill, he carved a figure, brilliantly, out of snow-white ivory, no mortal woman, and fell in love with his own creation. - Ovid, Metamorphoses*

# Pygmalion and Galatea



# Pygmalion and Galatea

- Pygmalion, who created Galatea
  - Creator and Creation Relationship
  - Perfection and Flaw
  - Anthropomorphization of Inanimate Objects
- Galatea, created by Pygmalion
  - Consciousness and Sentience
  - Ethical Creation
  - Transformative Power of Emotions
  - The Ideal vs. the Real



# Animated Artefacts

- *Hyakki Tsurezure Bukuro* (“The Illustrated Bag of One Hundred Haunted Artefacts” or “A Horde of One Hundred Haunted Housewares”).
- Hand-scrolls illustrated by Toriyama Sekien (1712 - 1788).

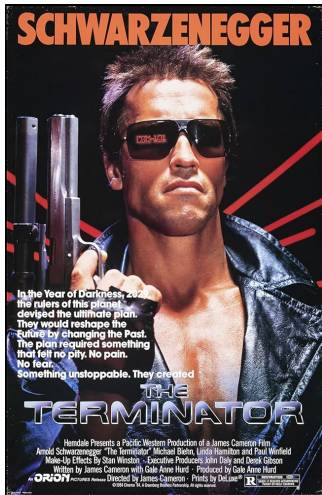


# Animated Artefacts

- *Koto-furunushi* illustrated by Toriyama Sekien (1712 - 1788).
- Said to be the music instrument of the legendary Emperor Keikō (71–130) of Japan.



# Idea Shift on Artificial Intelligence



- The Terminator (1984)



- The Matrix (1999)

# Idea Shift on Artificial Intelligence

- The Mass Effect Trilogy (2007) and Mass Effect: Legendary (2021)



# Idea Shift on Artificial intelligence



● Ex Machina (2014)



● Mass Effect: Andromeda (2017)

# Idea Shift on Artificial Intelligence



● Detroit: Become Human (2018)



● Wandavision (2021)

# Idea Shift on Artificial Intelligence

- ... *Organics don't know how we were created. Some say by chance. Some say by miracle: it's a mystery. But synthetics know we created them, and they know we are flawed. They are immortal, we are not. They see time as an illusion; we are trapped by its limitations. Above all, machines know the reasons why they were created. - Javik, Mass Effect Trilogy*

# What Now?

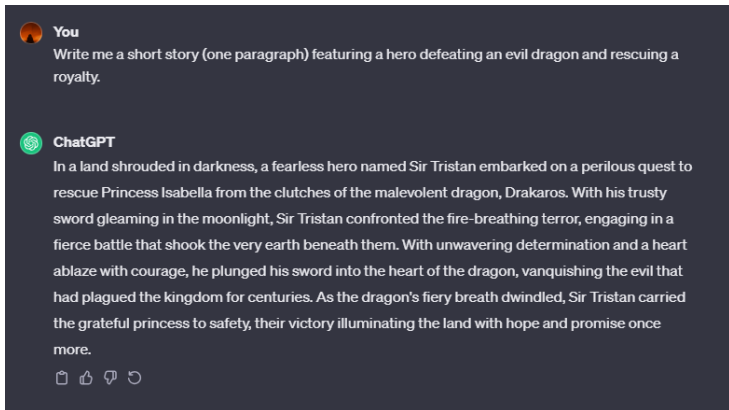


# ChatGPT

- Text Generation
  - Generate human-like text based on input prompts
- Answering Questions
  - Explain concepts and provide information on a wide range of topics
- Text Summarization
  - Summarize long articles, documents, or webpages into shorter, more concise versions
- Coding Assistance
  - Help with programming-related queries and code generation.
- Storytelling
  - Create and continue stories or narratives.
- What's the problem with these?

# ChatGPT: Generated Story

- Thoughts on this generated story?



**You**  
Write me a short story (one paragraph) featuring a hero defeating an evil dragon and rescuing a royalty.

**ChatGPT**  
In a land shrouded in darkness, a fearless hero named Sir Tristan embarked on a perilous quest to rescue Princess Isabella from the clutches of the malevolent dragon, Drakaros. With his trusty sword gleaming in the moonlight, Sir Tristan confronted the fire-breathing terror, engaging in a fierce battle that shook the very earth beneath them. With unwavering determination and a heart ablaze with courage, he plunged his sword into the heart of the dragon, vanquishing the evil that had plagued the kingdom for centuries. As the dragon's fiery breath dwindled, Sir Tristan carried the grateful princess to safety, their victory illuminating the land with hope and promise once more.

🗑️ 👍 🗨️ ↻

# Diffusion Model

- Architecture: DALL-E 2 uses a two-step process involving a diffusion model and a CLIP (Contrastive Language–Image Pretraining) model.
- The diffusion model generates images, and the CLIP model, trained to understand and rank the relevance of images in relation to text, guides the generation process.
- What's the problem with the results it generated, or just generative AI in general?
  - Biases in the training dataset!

# Diffusion Model



You

Provide a painting that describe this quote by Scott Fitzgerald: "Give me a hero and I will write you a tragedy."



DALL-E





Here are the paintings that capture the essence of the quote by Scott Fitzgerald: "Give me a hero and I will write you a tragedy." Each image portrays a heroic figure set against a backdrop of looming tragedy, reflecting the dramatic and emotional theme of the quote.




# Diffusion Model


- After some specifications in the prompt:

 **You**  
Can the paintings feature a woman instead of a man?  
< 2 / 2 >

 **DALL-E**







"GIVE ME A THE HERO  
AND I WILL WRITE A TRAGEDY"  
-SCOTT FITZGERALD

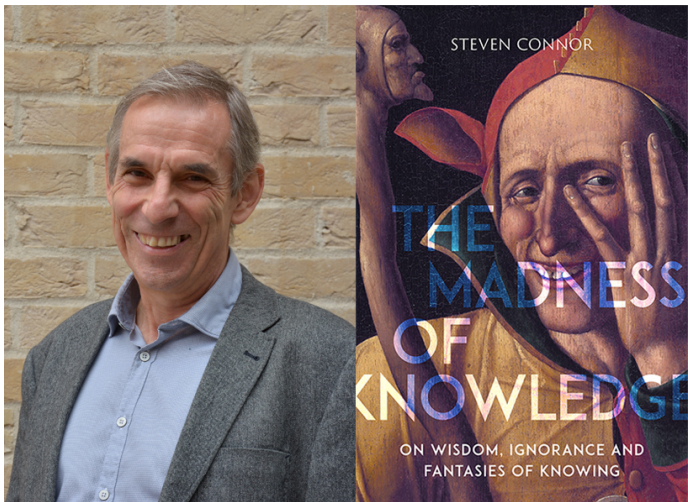


"GIVE ME AS A HERO AND  
I WILL WRITE A TRAGEDY"  
-SCOTT FITZGERALD

Here are the revised paintings, each featuring a heroic woman as the central figure, in line with Scott Fitzgerald's quote: "Give me a hero and I will write you a tragedy."

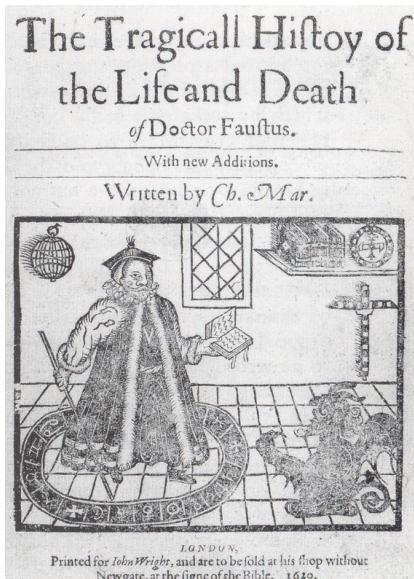
# On the Concept of “Learning”



# On the Concept of “Learning”

- **Fantasies of Knowing**
  - The author investigates fantasies of perfect knowledge, including the notion that total understanding is achievable.
- **Limits of Knowledge**
  - Connor discusses the ethical and existential implications of pursuing knowledge without bounds. He reflects on how the relentless drive for discovery can lead to ethical quandaries, personal madness, and even environmental degradation.
- **Wisdom and Madness**
  - The book considers the relationship between wisdom and knowledge, suggesting that true wisdom requires an awareness of the limits of one’s knowledge. It also delves into the idea that there is a fine line between the genius of discovery and what society considers madness.

# On the Concept of "Learning"

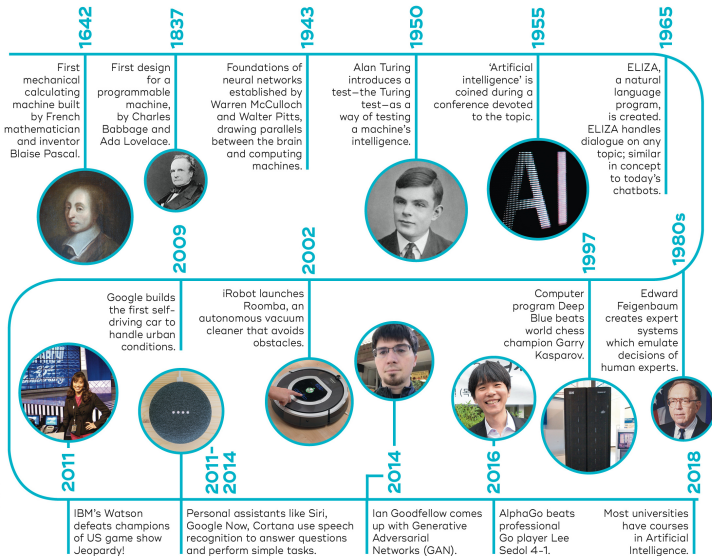




# What about machine learning?

By far we've only talked about human learning.

# History of Artificial Intelligence



# History of Neural Networks (High-Level)



**Warren McCulloch & Walter Pitts**, wrote a paper on how neurons might work; they modeled a simple neural network with electrical circuits.

**Nathaniel Rochester** from the IBM research laboratories led the first effort to simulate a neural network.

**John von Neumann** suggested imitating simple neuron functions by using telegraph relays or vacuum tubes.

STORY BY DATA

1943

1949

1950s

1956

1957

1958

## HISTORY OF NEURAL NETWORKS

1943-2019

**Donald Hebb** reinforced the concept of neurons in his book, *The Organization of Behavior*. It pointed out that neural pathways are strengthened each time they are used.

The **Dartmouth Summer Research Project** on Artificial Intelligence provided a boost to both artificial intelligence and neural networks.

**Frank Rosenblatt** began work on the Perceptron; the oldest neural network still in use today.

1982

1981

1969

1959

1982

**John Hopfield** presented a paper to the national Academy of Sciences. His approach to create useful devices; he was likeable, articulate, and charismatic.

Progress on neural network research halted due fear, unfulfilled claims, etc.

**Marvin Minsky & Seymour Papert** proved the Perceptron to be limited in their book, *Perceptrons*.

**Bernard Widrow & Marcian Hoff** of Stanford developed models they called ADALINE and MADALINE; the first neural network to be applied to a real world problem.

1982

1985

1997

1998

NOW

**US-Japan Joint Conference on Cooperative/Competitive Neural Networks**; Japan announced their Fifth-Generation effort resulted in US worrying about being left behind and restarted the funding in US.

American Institute of Physics began what has become an annual meeting - **Neural Networks for Computing**.

A recurrent neural network framework, LSTM was proposed by **Schmidhuber & Hochreiter**.

**Yann LeCun** published *Gradient-Based Learning Applied to Document Recognition*.

Neural networks discussions are prevalent; the future is here!

# History of Neural Networks (Low-Level)

