

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

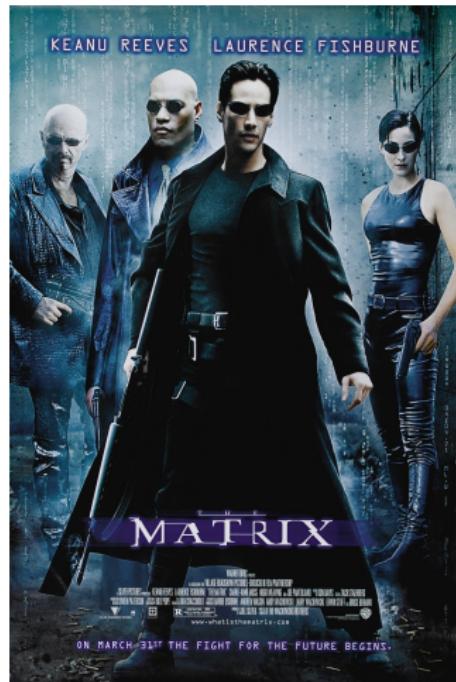
- *Koto-furunushi* illustrated by Toriyama Sekien (1712 - 1788).
- Said to be the music instrument of the legendary Emperor Keikō (71–130) of Japan.



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

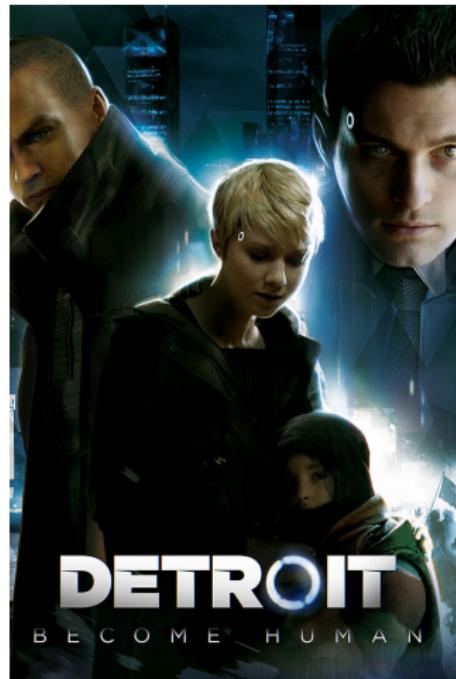


- Ex Machina (2014)



- 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?

The screenshot shows a conversation between a user and ChatGPT. The user, represented by a orange circular icon with a small tree, asks for a short story featuring a hero defeating a dragon and rescuing royalty. ChatGPT, represented by a green circular icon with a brain-like symbol, generates a detailed response. The response describes a hero named Sir Tristan embarking on a quest to rescue Princess Isabella from a malevolent dragon named Drakaros. The story details the fierce battle between Sir Tristan and the dragon, culminating in Sir Tristan's victory and the rescue of the princess.

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.

Reply Like Dislike 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 >

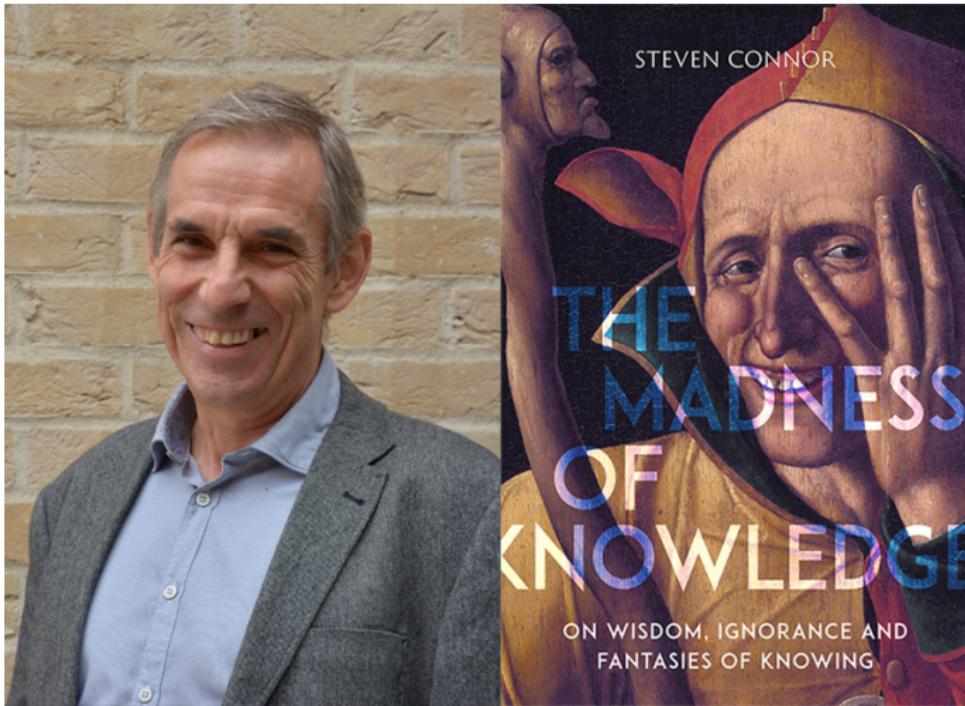
DALL-E

The image shows two side-by-side paintings generated by DALL-E. Both paintings feature a woman with curly hair, standing in a dramatic, fiery landscape with smoke and fire in the background. The woman is dressed in a dark, flowing garment. In the bottom left corner of each painting, there is a quote from Scott Fitzgerald: "GIVE ME A THE HERO AND I WIILL WRITE A TRAGEDY" followed by "-SCOTT FITZGERALD". The overall style is painterly and dramatic.

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."

□ ▲ □ □ □ □

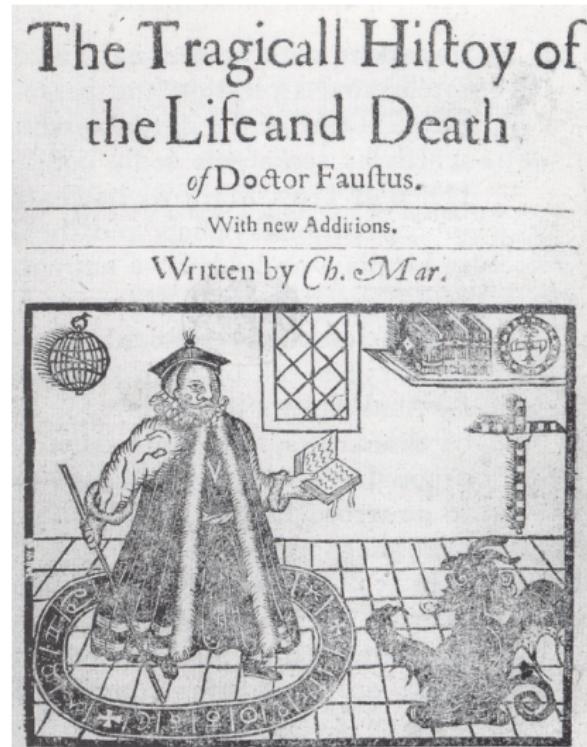
On the Concept of “Learning”



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”

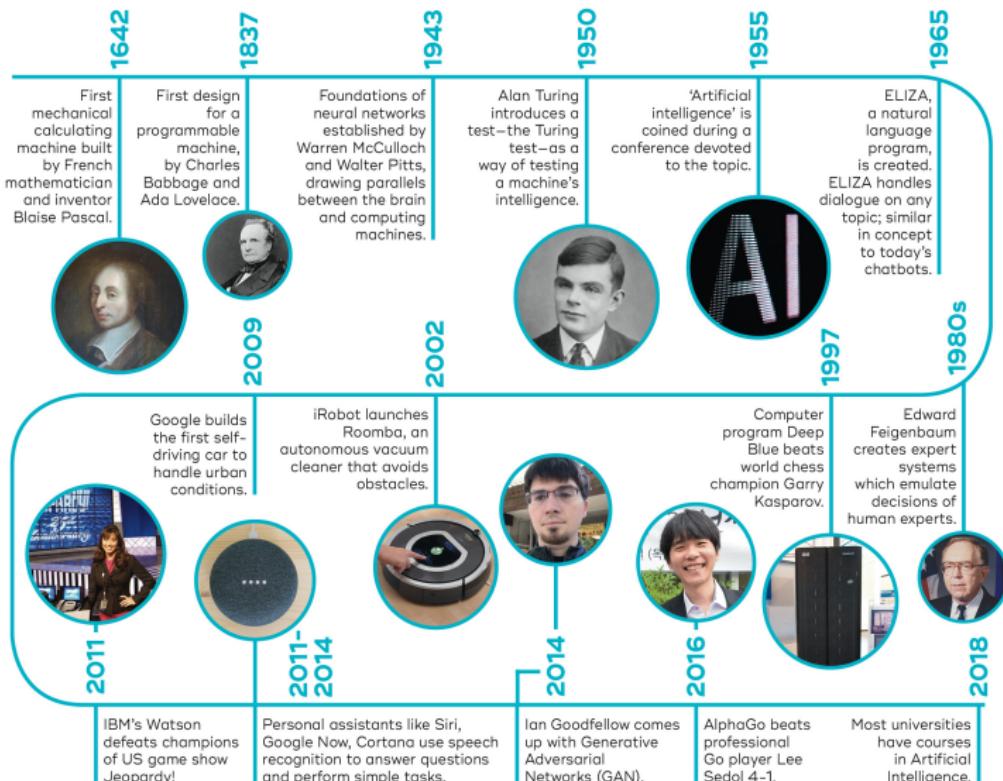


LONDON,
Printed for John Wright, and are to be sold at his shop without
Newgate, at the signe of the Bible. 1604.

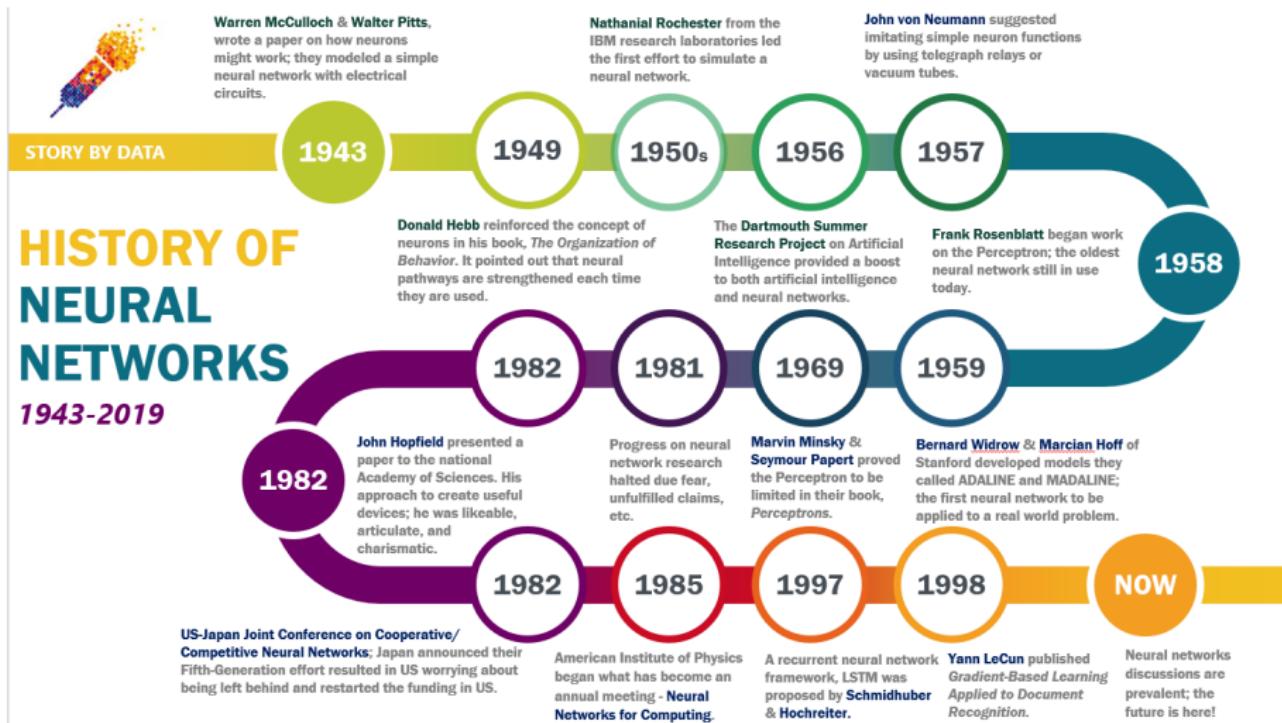
What about machine learning?

By far we've only talked about human learning.

History of Artificial Intelligence



History of Neural Networks (High-Level)



History of Neural Networks (Low-Level)

