AI FOR EVERYONE: EXPLORING THE POWER AND IMPACT OF ARTIFICIAL INTELLIGENCE

Dr. Clint Staley (C’80)

Principia Lifelong Learning Faculty

Former Principia College Computer Science Professor

Professor of STEM, UATX

CSC Professor Emeritus, Cal Poly SLO

Possible AI Topics

- **ChatGPT and its siblings**
- **Other important forms of AI**
- **How AI software “thinks”**
- History of AI development
- **Social and economic challenges of AI**
- **Deep questions about the nature of intelligence**
- Technical discussion of current AI research

ChatGPT and “LLM”s

- ChatGPT is one of a family of Large Language Models
  - Many LLMs available
  - ChatGPT (OpenAI), Claude (Anthropic), Grok (X), R1 (DeepSeek), Gemini (Google)
- Access is via simple, freemium websites
  - “Talk” in text-chat form with what seems like a very knowledgeable human, but is a program

Typical LLM Uses

- Comprehensive web search
- Summarizing and analyzing documents
- Generating or editing prose
- Tutorial discussions
- Brainstorming and recommendations

LLM Use “How To”

- It’s all about the prompt(s)
  - More than a question. Can include information, like documents, photos, etc.
  - LLM “knows” info only at the time of training, sometimes years earlier. Prompt gives current details.
  - Not just one prompt – engage in conversation. Entire conversation is “input” for each chat
  - Depth of remembered conversation varies by LLM – 1,000 words to 100,000
- LLM’s “hallucinate”, often very confidently
  - Check critical facts
  - Scan generated material for oddities

Example Uses

- Better web hunt – advice on furniture problem
  - Gives full description, various solutions, etc. (and it worked!)
- Document summary/analysis – IP agreement
  - Prompt includes entire document
  - Summarizes, points out issues, correctly evaluates
  - (But check for accuracy!)

Example Uses, cont..

- Editing prose – remove passive voice
  - Understands grammar, style
  - Example revision for PV is correct, succinct
- Brainstorming – name a cat
  - Sometimes offers silly suggestions with good ones
  - Will often come up with ideas you didn’t think of

Example Uses, cont..

- Write standard documents – reference letter
  - Prompt should include particular details, and general instructions
  - Can reuse prompt over and over w/ adjustments
  - Good at first drafts, but needs final polish (3rd paragraph, for instance)
- Tutorial Study – hierarchical Bayes' analysis
  - Gives good "textbook" summaries and explanations
  - Also response intuitively to questions
  - Wide ranging knowledge – like talking with universal graduate TA

Other AI

- Same principles of training and numerical patterns, but different designs and purposes.
- Diffusion Models
  - Generated movies in our intro
- Reinforcement Learning (RL)
  - Self-instructing. Requires automatically generable feedback.
  - Alpha Go example
- Physical AI
  - “Convolutional Neural Networks” or CNNs
  - Self-driving cars, robotics, face recognition

Example Uses – Other AI

- Artistic creativity – funny image
  - “Diffusion model”, e.g. DALL-E, Sora
  - Photos, short movie clips
  - May need feedback and adjustment, and may hallucinate
  - (Term “hallucination” originates with this type of AI)
- Generate audio of text **in your own voice**
  - Very convincing
  - Note that this means voices are not representative e.g. on phone calls

Higher “Reasoning” in LLMs

- Intuitively understanding higher math (math tutorial example)
- Writing and Revising a Story
  - Understands the context of the story
  - Summarizes; changes style
- Analyzing humorous images
  - “Gets” the humor
- Theory of Mind
  - Can empathize and visualize what’s in people’s minds

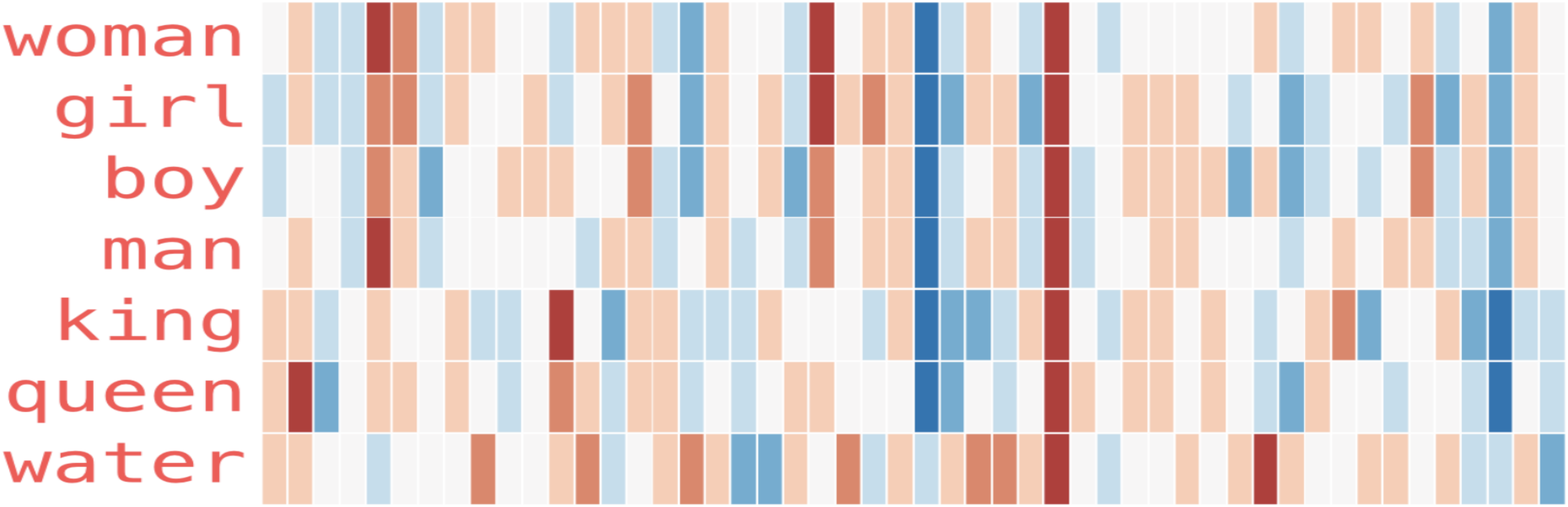
An LLM Is *Not*

- An improved search engine
  - Is able to web-search when useful, but…
  - … can work unplugged from the internet
- Hand-crafted code
  - No programming specific to the questions/answers
  - Could be written in one page of code – **12 equations**
- The whole modern AI story
  - Important other areas are making major advances too, e.g. “reinforcement learning”, computer vision

An LLM *Is*

- A sophisticated mathematical engine
  - 12 equations, but on “matrices” – huge grids of numbers
  - From 10 billion to over 1 trillion numbers
  - The “intelligence” is in those numbers (aka weights or parameters)
- Self-organizing!
  - The weights are “trained” by incremental adjustment
    - First, to match known texts
    - Then to adjust to human feedback given by RLHF
- Humans design the training process and the equations, but not the weights

Word/Idea “Embedding”



- This is how an LLM represents words, phrases and ideas
- Each word has a number sequence – an “embedding”
- But, with 12,800 numbers per word – many conceptual dimensions

How An LLM “Thinks”

- Starts with word embeddings representing words in your prompt and the current conversation.
- Combines words into higher 12,800-number patterns representing phrases, then sentences, etc.
- Predicts next “thought” as a number pattern and chooses closest next word.
- Repeat, word by word...

The Weights *are* the “Intelligence”

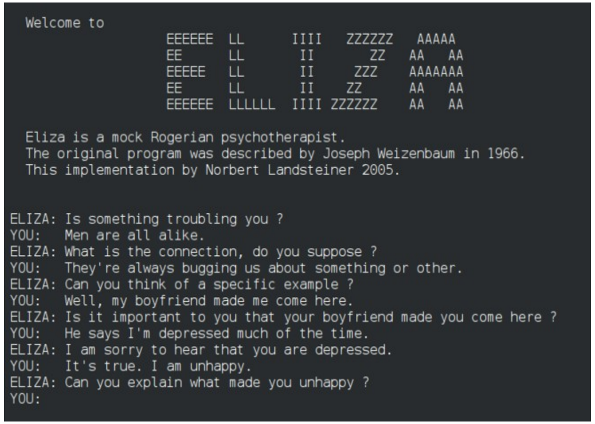
- All “reasoning” is by word/concept embeddings.
- And we (mostly) *don’t know* what each weight does.
- Could learn to converse in a language none of its programmers understand.
- Are world-models, theory of mind, etc. spontaneously emerging from the training?

Is an LLM Intelligent?

- “Yes” arguments
  - Vast possibilities in 1 trillion weights. Printed out they’d be 4 million large books.
  - Behavior suggests world-models, theory of mind, etc.
  - (But, we really don’t know since we’re not sure how the weights work in detail)
- “No” arguments
  - Just mindless numbers and 12 equations
  - Just predicting the next word.
  - (But, it’s improving *fast*, a human brain is just “mindless” neurons and synapses, and *how* is it picking that next word so intelligently?)

AI History 1950s - 1960s

- (1956) “Artificial Intelligence” first used as a term
- (1956) First good “theorem prover” – **symbolic AI**
- (1965) ELIZA basic “chatbot” in 200 lines of code
- (1960s) Hopfield, Minsky and others do early work in **“neural networks”**
- (1967) Minsky: “Within a generation … the problem of creating artificial intelligence will substantially be solved.”
- (1969) Minsky and Papert publish “Perceptrons”
  - Proves one-layer (all they knew how to train) neural networks very limited and triggers first **“AI winter”**



```
  Welcome to
                    EEEEEE  LL      IIII   ZZZZZZ   AAAAA
                    EE      LL       II       ZZ   AA   AA
                    EEEEE   LL       II     ZZZ    AAAAAAA
                    EE      LL       II    ZZ      AA   AA
                    EEEEEE  LLLLLL  IIII ZZZZZZ    AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

AI History 1970s-1990s

- (1970) Seppo Linnainmaa quietly publishes “backpropagation” algorithm, solving the multilayer training problem. *No one notices.*
- (1970s and 80s): Direct coding of logic in “expert systems”
  - Turns out to be limited and fragile
  - Vaunted “5th generation” computing fizzles
- (1986) Backpropagation algorithm (re)discovered.
  - Neural nets are back.
  - But training networks > 3 layers proves very hard. (Modern AI takes hundreds.)
- After brief burst of progress, a **second “AI Winter”**

AI History 2012-2024

- Four important advances lay ground for new neural network growth
  - more available training data via internet
  - computing power via GPUs
  - better neural network design: “activation functions” and architectures (neural net => deep learning)
  - vibrant and very intensive research community
- (2012) Neural networks, now “deep learning networks” are back, again!
- (2017) “Attention is all you Need” paper describes Transformers
- (2020-2023) Steady LLM improvement arrives at GPT-4

AI History – Big Ideas

- Early investigation of major ideas
  - Neural networks and “chatbots” since 1960s
- Alternating optimism and AI winters
  - Two major plateaus in progress
- Symbolic logic vs neural networks
  - The “bitter” (to some researchers) lesson – can’t program AI directly
- What’s different this time?
  - GPUs, massive training data, better algorithms, vibrant research community

Recent AI Events

- Nobel Prizes
  - Physics to John Hopfield and Geoffrey Hinton
  - Chemistry to the developers of Alpha Fold
- Improved LLM Architectures
  - MoE – “Mixture of Experts”
    - DeepSeek example
  - CoT – “Chain of Thought” – LLMs that reason
- Improvements in diffusion models
  - Sora movie generation

AI Events: late 2020s (?)

- Ongoing automation of white collar work
  - Writing, coding, art, music, law, healthcare, customer service, language translation, entertainment, etc.
  - My own experience: It's like a competent, widely read, and blindingly fast junior engineer.
  - May double or triple productivity of experienced professionals.
- Physical AI
  - AI with vision, movement, hearing
  - Robotics, self-driving cars, manufacturing, construction, etc.
- Scientific advances using AI, especially RL

Near Term Issues to Consider

- Legal questions of copyright and liability
- “Alignment” of AI with human values
- AI enables experienced professionals but replaces beginners creating “apprentice gap”
- Technological unemployment
  - Not like going from farm to factory to simple white-collar work
  - UBI might provide subsistence, but doesn’t provide purpose
- Which nation will lead? – Manhattan project of the 2020s

AI Future – Where is it heading?

- “… heavier than air flying machines are impossible”
  - Lord Kelvin, 1895
- Alchemistic Empiricism
  - Analogy of Chemistry in 1825
  - Results in fits and starts as we try things out by guess-and-by-golly
- Start with Nature, End with Design
  - Planes vs birds
  - Deep Learning vs brains : Geoff Hinton’s warning

AI Events: 2030s (??)

- Option A: AI progress will “plateau”
  - We are reaching limits on compute and on data
  - It’s plateaued before
- Option B: Continued progress reaching “AGI” or “superintelligence”
  - Unprecedented research community and infrastructure
  - Continued improvement in design and training procedures
  - Hardware is rapidly improving
- 5 years ago, many experts thought “A”. Now many think “B”.

Long Term Questions

- What can AI teach us about the nature of thought and intelligence?
- What if it becomes more intelligent than us?
- Hard Problem of Consciousness
- “Who is my neighbor?”

Advice to Students

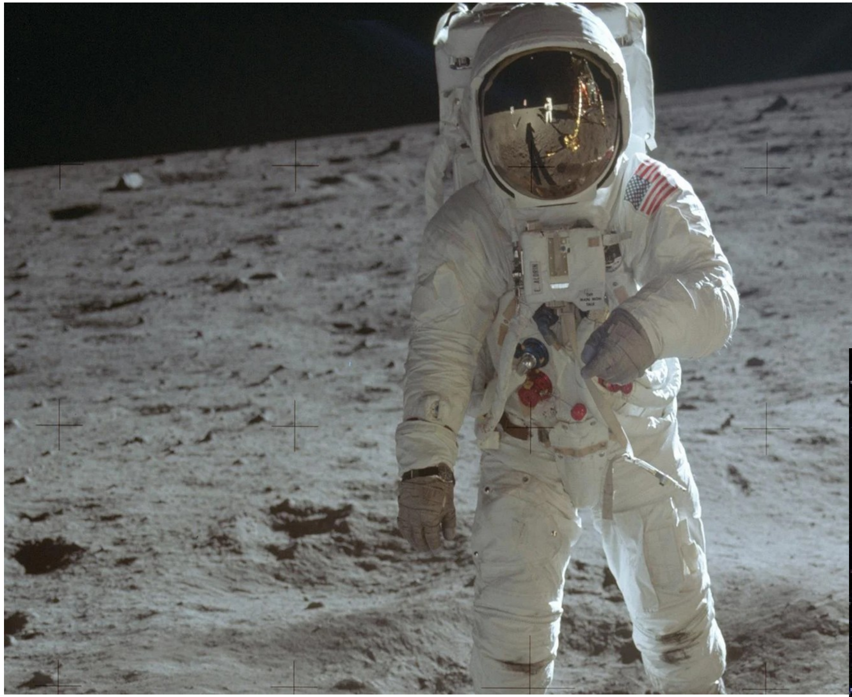
- Learn to create without AI first
  - Can’t be “in charge” if you are dependent
  - Gets you closer to content, like doing your own calculations
- Then, create using AI
  - Leverage AI to create more than you can alone
  - Use AI as “cointelligence” partner from whom you can learn as well
  - This is the workflow of the future

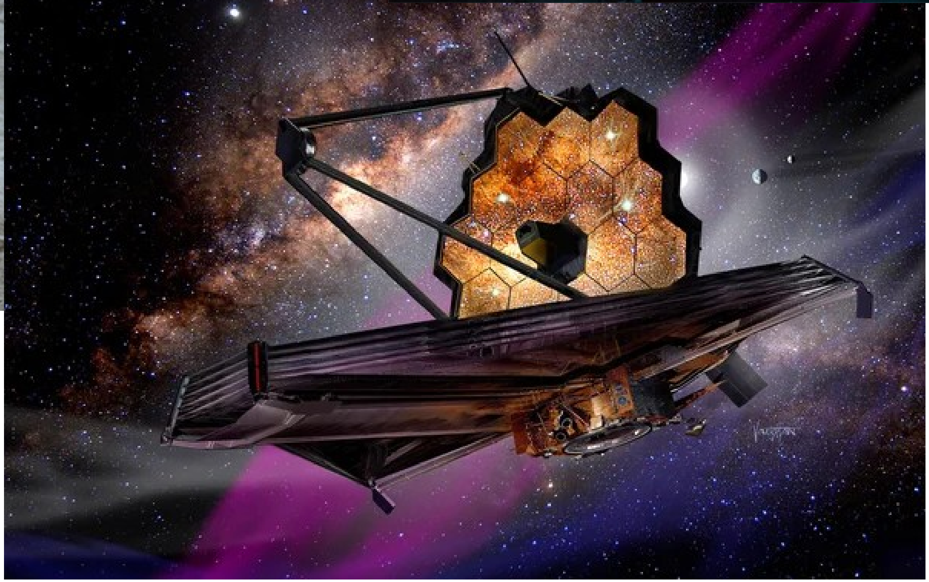
Future Predictions – Dystopian SF

- Premise: Technology is a power unto itself
- (1984) Blade Runner
  - LA is stygian hellscape in 2020s
- (1984) Terminator
  - LA is nuclear-blasted wreck with robots killing humans in 2020s
- (1970) Colossus
  - AI computer becomes world dictator in late 1970s

Future Predictions – CS Perspective

- Premise: Technology is Mind’s creativity unfolding.
- “The mariner will have dominion over the atmosphere and the great deep .. the astronomer will no longer look up to the stars, – he will look out from them..” Science and Health p125: 25-29





Christian Science Perspective

- *“There is no life, truth, intelligence, nor substance in matter. All is infinite Mind and its infinite manifestation…” S&H p 468*
- *“Intelligence is omniscience, omnipresence, and omnipotence. It is the primal and eternal quality of infinite Mind...” S&H p 469*

Current Research – Optimization

- Make using AI less expensive after training
  - Reduce power consumption
  - Allow LLMs to run on edge devices like phones
- Not all weights are equally valuable; groups of weights can be compressed
- Precision of the weights can be looser, even down to one bit
- “Speculative Decoding”
  - Small “draft” LLM model does initial cut, confirmed by stronger model
- Knowledge Distillation (KD)
  - Train small LLM from “smart” larger one that offers nuance

Current Research – Intelligence

- Longer context windows
  - Squared complexity – requires creative selectivity in pairwise comparisons
  - Reasoning models
  - Architecture improvements: RoPE, better activation functions
  - Improved generalization
    - More data examples
    - RL-based training, especially in physical AI, coding, etc.
    - Double descent pattern and “grokking”

Current Research – Explainability

- Analyzing embeddings for concept patterns
  - Sparse autoencoders, and “golden gate bridge”
- Mechanistic Interpretability (MI)
  - Reverse engineering weight subnets “circuits”
- High Math Approaches
  - Mutual Information Theory
  - Sheaf Theory
  - Analysis of high-dimensional optimization surfaces, e.g. singular learning theory (SLT)