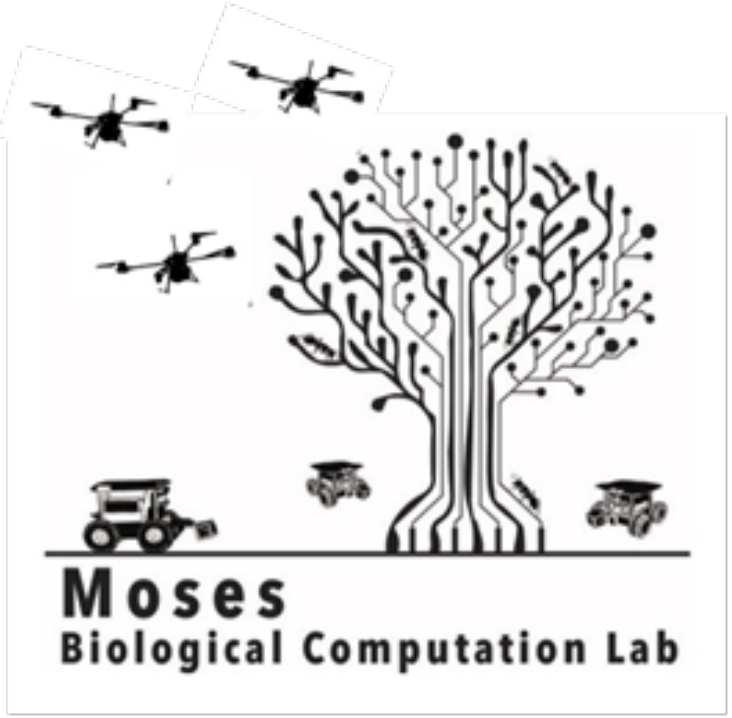
# AI (R)evolution:
# How did we get here
# And where are we going?

Melanie Moses
UNM Professor of Computer Science
External Faculty, Santa Fe Institute
melaniem**unm.edu

# My perspective

- History
  - Began studying AI in 1989 at Stanford
  - Worked in computer security in the 1990s
  - Ph.D. UNM Biology 2005
- Current Research
  - Interdisciplinary, Complex Adaptive Systems
  - Roboticist: VolCAN
  - Computational Biologist: SIMCoV
- Concerns
  - Encoded bias, surveillance & concentrated power
  - AI that can code faster than students
  - How humans will co-evolve with AI



# My perspective

- History
  - Began studying AI in 1989 at Stanford
  - Worked in computer security in the 1990s
  - Ph.D. UNM Biology 2005
- Current Research
  - Interdisciplinary, Complex Adaptive Systems
  - Roboticist: VolCAN
  - Computational Biologist: SIMCoV
- Concerns
  - Encoded bias, surveillance & concentrated power
  - AI that can code faster than students
  - How humans will co-evolve with AI

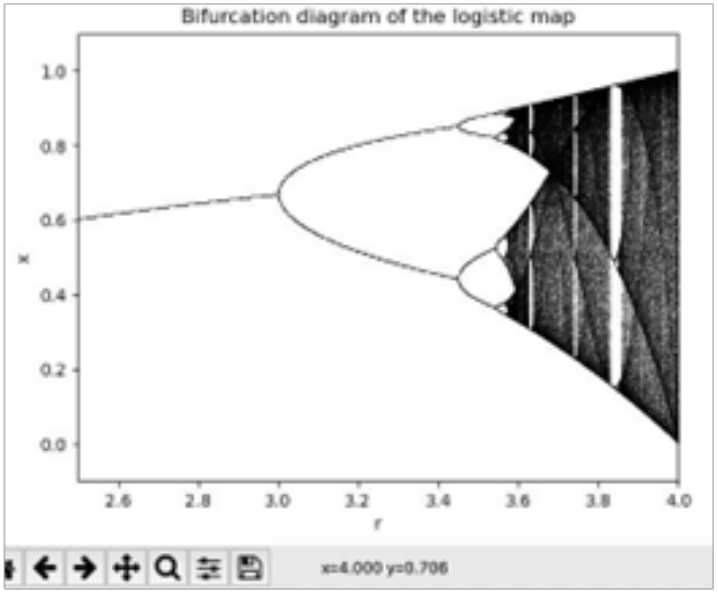
# My perspective

- History
  - Began studying AI in 1989 at Stanford
  - Worked in computer security in the 1990s
  - Ph.D. UNM Biology 2005
- Current Research
  - Interdisciplinary, Complex Adaptive Systems
  - Roboticist: VolCAN
  - Computational Biologist: SIMCoV
- Concerns
  - Encoded bias, surveillance & concentrated power
  - AI that can code faster than students
  - How humans will co-evolve with AI

www.dailymaverick.co.za/article/2021-04-16-this-weekend-were-watching-coded-bias/


https://www.codedbias.com/

# ChatGPT: do my students' homework




Here's a Python code using `matplotlib` to display the logistic map:

python

Copy code

```
import numpy as np
import matplotlib.pyplot as plt

def logistic_map(r, x):
    return r * x * (1 - x)

def plot_logistic_map():
    r = np.linspace(2.5, 4.0, 10000)  # Values of r
    x = 1e-5 * np.ones_like(r)  # Initial condition for x

    iterations = 1000
    last = 100  # We'll only display the last few iterations to see the ste

    for i in range(iterations):
        x = logistic_map(r, x)
        # Only display the bifurcation diagram for the last few iterations
        if i >= (iterations - last):
            plt.plot(r, x, ',k', alpha=0.25)

    plt.xlim(2.5, 4)
    plt.ylim(-0.1, 1.1)
    plt.xlabel('r')
    plt.ylabel('x')
    plt.title('Bifurcation diagram of the logistic map')
    plt.show()

plot_logistic_map()
```

To see the bifurcation diagram:

1. Install the necessary libraries: `pip install numpy matplotlib`.
2. Run the above code.

Regenerate

# My perspective

- History
  - Began studying AI in 1989 at Stanford
  - Worked in computer security in the 1990s
  - Ph.D. UNM Biology 2005
- Current Research
  - Interdisciplinary, Complex Adaptive Systems
  - Roboticist: VolCAN
  - Computational Biologist: SIMCoV
- Concerns
  - Encoded bias, surveillance & concentrated power
  - AI that can code faster than students
  - How humans will co-evolve with AI

Generate DALL-E prompts …

ChatGPT: A market scene in 2150 where humans buy tech enhancements for their bodies, while robots buy organic parts to feel more human.

# My perspective

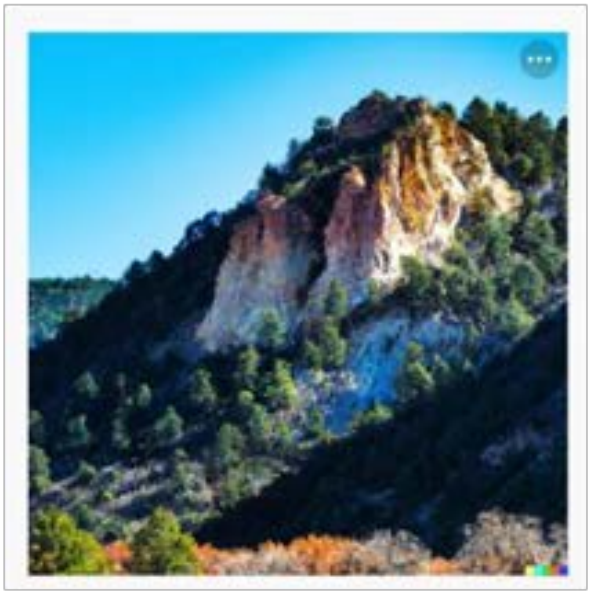
- History
  - Began studying AI in 1989 at Stanford
  - Worked in computer security in the 1990s
  - Ph.D. UNM Biology 2005
- Current Research
  - Interdisciplinary, Complex Adaptive Systems
  - Roboticist: VolCAN
  - Computational Biologist: SIMCoV
- Concerns
  - Encoded bias, surveillance & concentrated power
  - AI that can code faster than students
  - How humans will co-evolve with AI

Generate DALL-E prompts …

ChatGPT: A human and robot duo exploring alien landscapes with the guidance of an AI navigator.

# What I'll talk about today

- How does AI work?
- What can it do?
  - The good, the bad & the ugly
- How AI might evolve from here
- What you (and we all) can do



Note: Images generated by DALL-E 3
Text aided by GPT-4

# Interdisciplinary Working Group on Algorithmic Justice

Melanie Moses CS, UNM / SR

Cris Moore SR

Kathy Powers Poli Sci, UNM/ SFI

Alfred Mathewson Law, UNM

Sonia Rankin Law, UNM

Mirta Galesic SR

Josh Garland Arizona State

Matthew Fricke CS, UNM

Gabe Sanchez Poli Sci, UNM

Tina Eliassi-Rad CS, NEU /SFI

Mahzarin Banaji Psych, Harvard/SFI

Trilce Estrada CS, UNM

Nadiyah Humber Law, UConn

# How does AI work? Traditional algorithms vs. AI

Traditional computer programs are like recipes: step by step instructions to achieve a specified outcome.

A simple cake recipe

1. Preheat oven to 350 degrees F
2. Mix 2 cups sugar and 4 sticks butter
3. Add 8 eggs
4. Mix in 3 cups flour

...

Bake in greased pan for 50 minutes

http://www.mykitchenintherockies.com/high-altitude-baking/

# How does AI work? Traditional algorithms vs. AI

Traditional computer programs are like recipes: step by step instructions to achieve a specified outcome.

Traditional algorithms can fail, but we usually understand why

Generative AI is different: we don't know how it makes predictions

A simple cake recipe

1. Preheat oven to 350 degrees F
2. Mix 2 cups sugar and 4 sticks butter
3. Add 8 eggs
4. Mix in 3 cups flour

...

Bake in greased pan for 50 minutes

Unless you are above 5000 feet

Context matters!

http://www.mykitchenintherockies.com/high-altitude-baking/

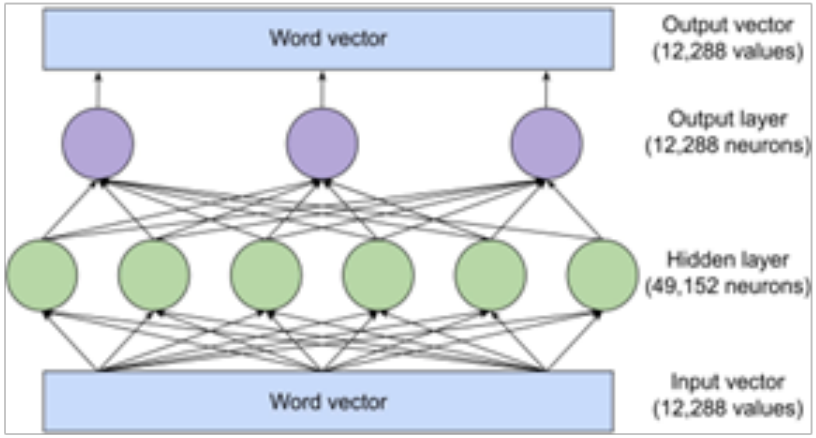
# How does AI work? Large Language Models (ChatGPT)

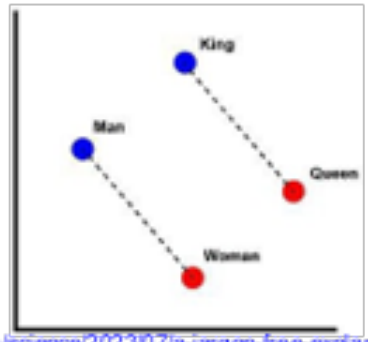
They use neural networks to find patterns & make predictions

Billions of nodes trained on billions of texts predict what word is likely to come next

Words are vectors (a list of 1000's of numbers) grouped by similarity

Neural networks are pattern matching machines




**Paris** - **France**
+ **Germany** = **Berlin**



https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/

# How does AI work? Large Language Models (ChatGPT)


LLMs group similar words together

https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/

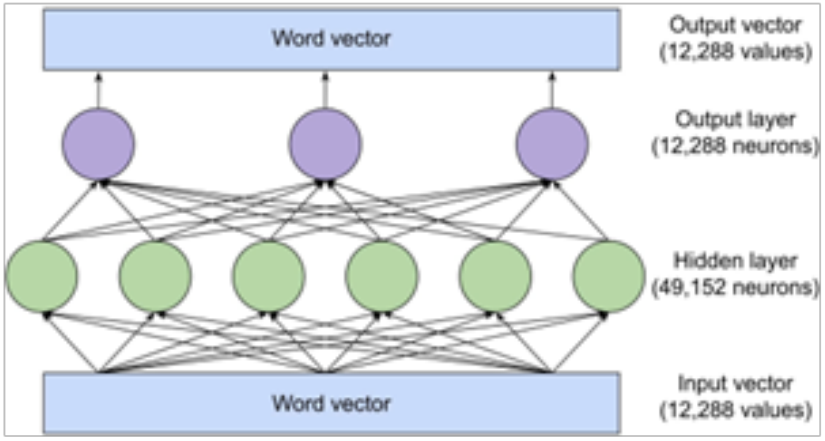
# LLMs learn by predicting what word comes next

Roses are red, violets are xxx

Cherries are red, grapes are xxx

Green means go, red means xxx

Hands have fingers, feet have xxx



# This is tricky!

Transformers: "attention mechanisms" disambiguate multiple possible meanings:

Bank: so many meanings!

Pronouns: The doctor gave the patient her stethoscope.

# This is tricky!

Transformers: "attention mechanisms" disambiguate multiple possible meanings:

Bank: so many meanings!

Pronouns: The doctor gave the patient her stethoscope.

"The University of New Mexico"

"Queen Bey" vs Queen Bee

## How does does disambiguation work?

**25 page paper!!**
Analyzed how GPT-2 predicted the next word for the passage

> "When Mary and John went to the store, John gave a drink to" …

## How does does disambiguation work? Hard to say!

25 page paper analyzes how GPT-2 predicted the next word for the passage

> "When Mary and John went to the store, John gave a drink to" … "Mary"
>
> but…
>
> "When Mary and John went to the restaurant, John gave his keys to," … "the valet."

Write a haiku about the promise of AI

Silent mind, awake,
AI's promise, boundless skies,
Humanity's guide.

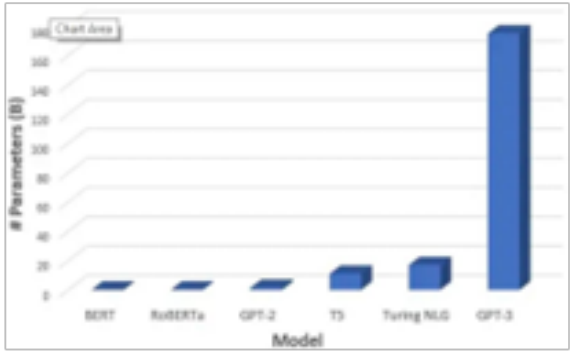
https://chat.chatgptdemo.net/

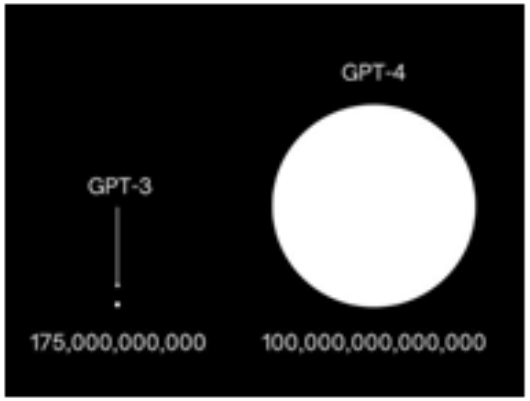
## Why did AI suddenly become so much better?

500 billion words

**300 billion trillion calculations** to train GPT-3.

People only hear 10 million words by age 10.



**Max Tegmark**
@tegmark

You know that AI is taking off when...Microsoft plans to power data centers with nuclear reactors:
jobs.careers.microsoft.com/global/en/job/...

We're looking for a Principal Program Manager, Nuclear Technology, who will be responsible for maturing and implementing a global Small Modular Reactor (SMR) and microreactor energy strategy.

This senior position is tasked with leading the technical assessment for the integration of SMR and microreactors to power the datacenters that the Microsoft Cloud and AI reside on.

# Scale is all you need?

Moore's Law:
**1970:1000 transistors → 2020: 50 BILLION transistors**

"Intelligence" at scale

- Climate change[1]
- Little context, many errors
- Large corporations dominate

Models trained with larger dataset more likely to classify Black faces as non-human or criminal[2]

Encodes the worst of the past & projects to the future


1 On the Dangers of Stochastic Parrots, Bender et al 2021
2 Birhane et al preprint 2023; arxiv.org/abs/2306.13141

# Human feedback is still required to train ChatGPT

TIME

BUSINESS + TECHNOLOGY

## Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic



*Content warning: this story contains descriptions of sexual abuse*

## Why did AI suddenly get so much more powerful?

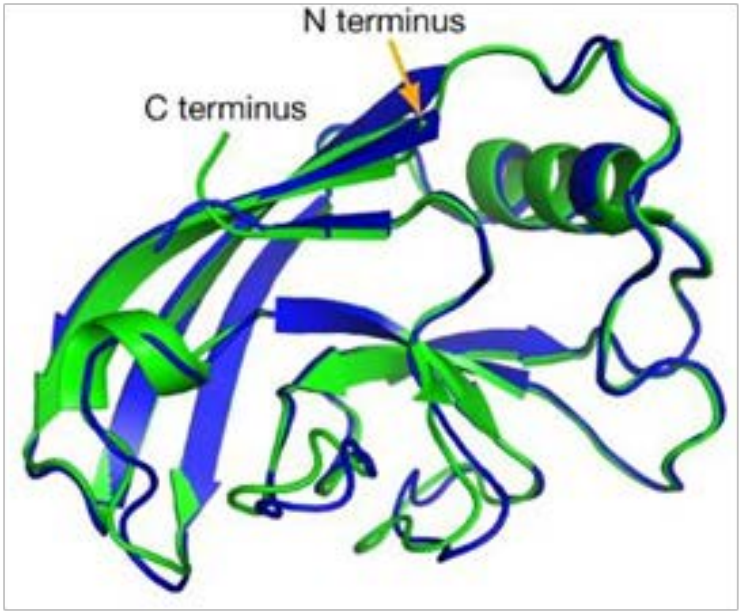
- Scale
- A few clever tricks in neural networks
  - Encoders, Transformers
- Human feedback
  - Low paid workers in Africa and the Global South

## Will it continue?

- Engaging users provides more feedback
- A dynamic ecosystem produces multimodal AI, Bing GPT4 + DALL-E 3
- Competitive acceleration in the market exponential growth & evolution
  - **Meta's Open-ish** Code LLama is smaller than its predecessors
- **Will LLM's learn to improve themselves?**

# AI: The Good

- Alpha-fold: Predicts protein structure from DNA sequences. Fundamental question in Biology.
  - 5 years of Ph.D. research done in minutes
- Applications in medicine: radiology, drug discovery, disease diagnosis, vaccine design
- Applications in basic science: New materials to store energy; minimize plastic waste; perhaps fundamental breakthroughs in physics & computation
- Robotics and autonomous driving (good?)




Highly accurate protein structure prediction with AlphaFold, *Nature* 2021

# AI: The Bad

The Problem

- Bias, errors & bad actors
- Criminals, corporations, governments & regular people can misuse AI.
- Misinformation will proliferate even faster



Fox News
NYC bans AI tool ChatGPT in schools ...

States' Automated Systems Are Trapping Citizens in Bureaucratic Nightmares With Their Lives on the Line



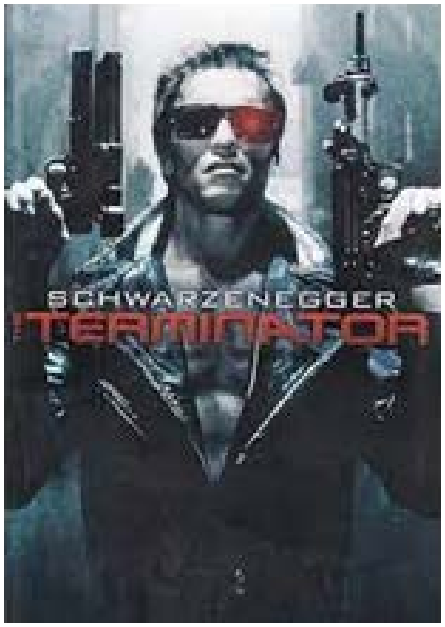
ABC News

Experts warn of rise in scammers using AI to mimic voices of loved ones in distress

# AI: The Ugly

- Concentrated Corporate Power
- Privacy & Surveillance
- Discrimination encoded & projected into the future
- Loss of jobs
- Rights of creatives
- We trust algorithms too much
  - Example: kidney algorithm

Not The Problem




(at least not yet)

# AI: Where are we going?

- Exponential growth of "intelligence". We do not know where that will lead.
- Restructuring of work; economic inequality; concentrated wealth and power
- "Alignment" of AI with "human values".

  Politicians: how good are we at determining what human values to align to?

Certainly! Here's a prompt for DALL-E-3 that aims to visualize AI alignment with human values:

"An AI and human shaking hands against a backdrop of a harmonious world, symbolizing a perfect blend of technology and humanity, with icons of peace, empathy, ethics, and collaboration surrounding them."

# How will AI evolve from here?

Modern AI is a tool that interacts with us using human language. It will use language to interact with our other tools.

Explosion in the generative AI ecosystem - competition for best models,integration, ease of use

It will be guided by market forces (making companies money) and user satisfaction (giving people tools they want to use) unless someone (you) regulate, guide, restrain, incentivize

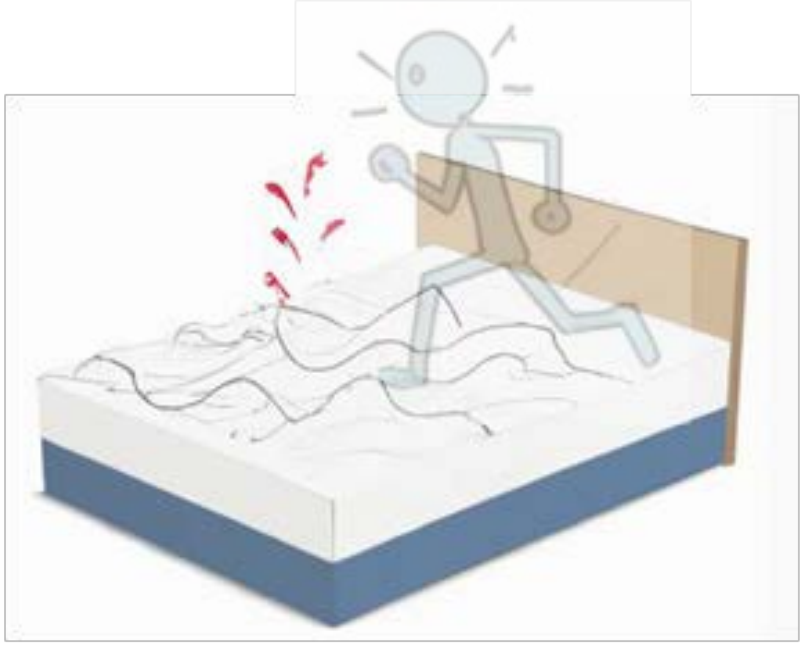
# What can you do? Now:

- **Require transparency**
  - Particularly in public sector (lending, criminal justice, housing) and any domain impacting civil rights
  - Humans should know what is created by AI
- **Regulate:**
  - misinformation, particularly in politics
  - corporate and government concentration of power, surveillance; right to appeal
- **Educate:** AI literacy for everyone: public, K-12, universities
  - Emphasize for targeted & vulnerable groups
- **Research**
  - Understand the social impact of AI
  - Develop AI designed for ethical use in the public interest
  - Mitigate & guard against harms from AI

# What can you (we) do? Long term

**Prepare for a marathon on a changing landscape:** Regulation & incentivization will need to adapt as AI and citizens evolve

**Engage the public in these decisions.** AI has the potential to transform society. It does so using input from (almost) all of us. It is the responsibility of all of us to shape the future.



Mashup: Running on a waterbed

# A few helpful resources

- How does AI work?

  https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/

- AI bias and surveillance

  https://www.codedbias.com

  https://www.rollingstone.com/culture/culture-features/women-warnings-ai-danger-risk-before-chatgpt-1234804367/

- Long term (existential, perhaps overblown) risks

  Impact Theory podcast interview with Mo Gawdat

  https://open.spotify.com/episode/2CiP3A1ZTcVlaNjFLURiox


UNM/SFI Algorithmic Justice
https://www.santafe.edu/research/projects/algorithmic-justice

Melanie Moses, Prof. of CS at UNM
melaniem**unm.edu