AI: The Need for an Inclusive Conversation

Arthur “Barney” Maccabe
Executive Director
Institute for Computation and Data-Enabled Insight
The University of Arizona

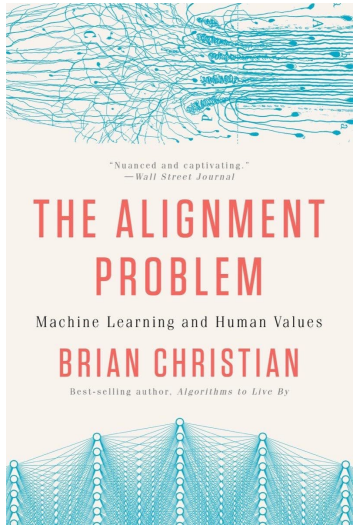
Coding Kindness: Regulatory Approaches to Technology and Compassion

Institute for Computation and Data-Enabled Insight

- ICDI: Established in March 2022 at the University of Arizona
  - https://datainsight.arizona.edu
- Amplifying existing and growing new activities in data and computation
- Growing connections to communities in Arizona

THE UNIVERSITY OF ARIZONA
Institute for Computation
& Data-Enabled Insight

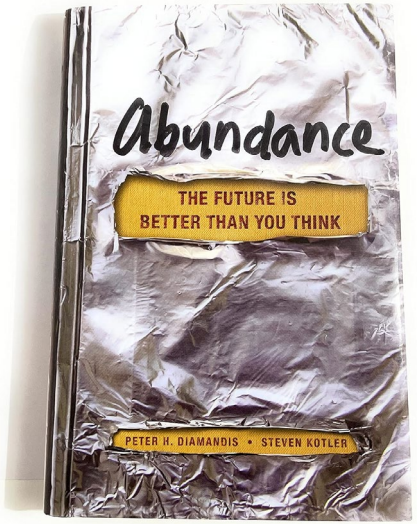
Presentation Summary



- AI is both exciting and terrifying
  - Opportunities are unbounded,
  - So is the potential for societal transformation
- Everyone needs to be engaged in the conversation
  - You don't need to be an expert in AI to participate
  - This is really about the society we want to have
- New rules, or re-interpretation of old rules?
- What does it mean to be human?

Exciting

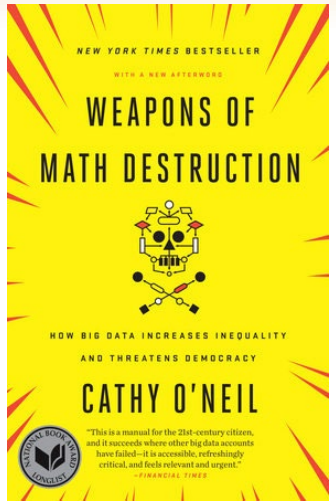
- Easy access to vast amounts of information
- Personalized education/learning
- Improved productivity*
- Compensating for differing abilities
- New kinds of jobs, e.g., prompt engineering

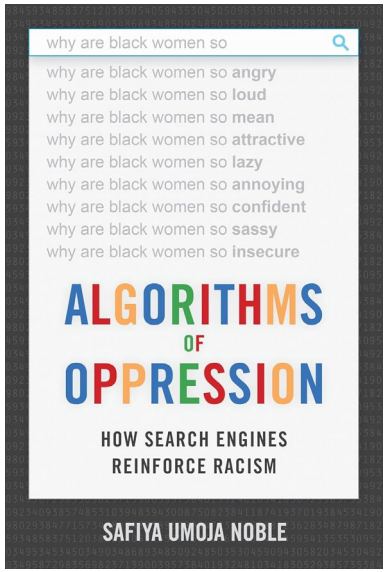


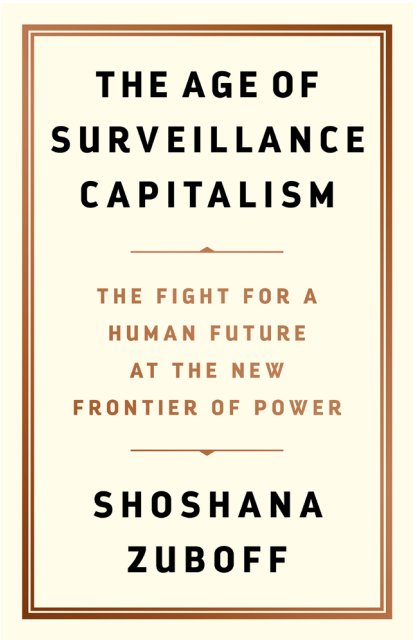
* I used ChatGPT-4 in developing this talk

Terrifying

- Job displacement
- Bias
- Data Privacy & Sovereignty
- Superintelligence







Job Displacement

Like many technologies of the past, AI is poised to displace many jobs.

Chief Justice Roberts' Year-End Report on the Federal Judiciary for 2023 focused on technology

"AI obviously has great potential to dramatically increase access to key information for lawyers and non-lawyers alike. But just as obviously it risks invading privacy interests and **dehumanizing** the law."

"Machines cannot fully replace key actors in court. Judges, for example, measure the sincerity of a defendant's allocution at sentencing. Nuance matters: Much can turn on a shaking hand, a quivering voice, a change of inflection, a bead of sweat, a moment's hesitation, a fleeting break in eye contact. And most people still trust humans more than machines to perceive and draw the right inferences from these clues."

Emphasis added for this presentation

https://www.supremecourt.gov/publicinfo/year-end/year-endreports.aspx

Chief Justice's Year-End Report on the Federal Judiciary (previous years for context)

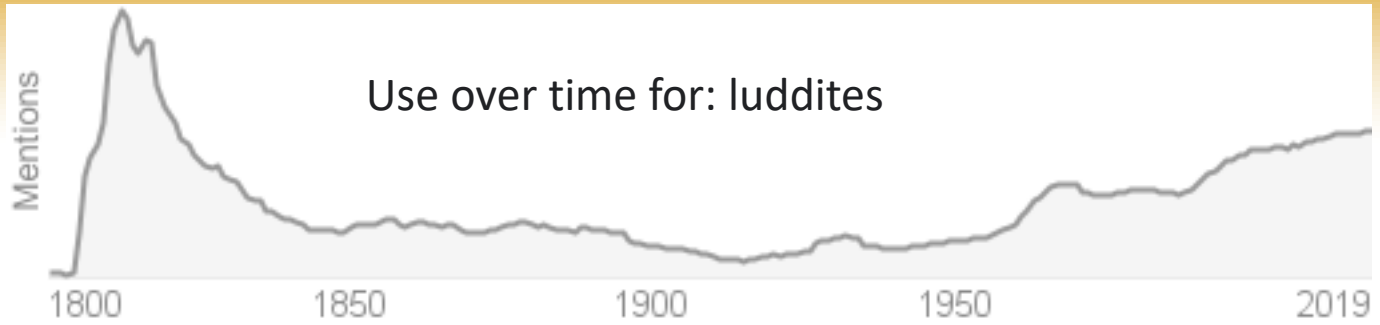
- 2020 Pandemic response by the judiciary
- 2021 Independence of the judiciary

  "The Congress of his [Chief Justice Taft] era appreciated the Judiciary's need for independence in our system of separate and co-equal branches, and it provided a sound structure for self-governance."
- 2022 Judicial security

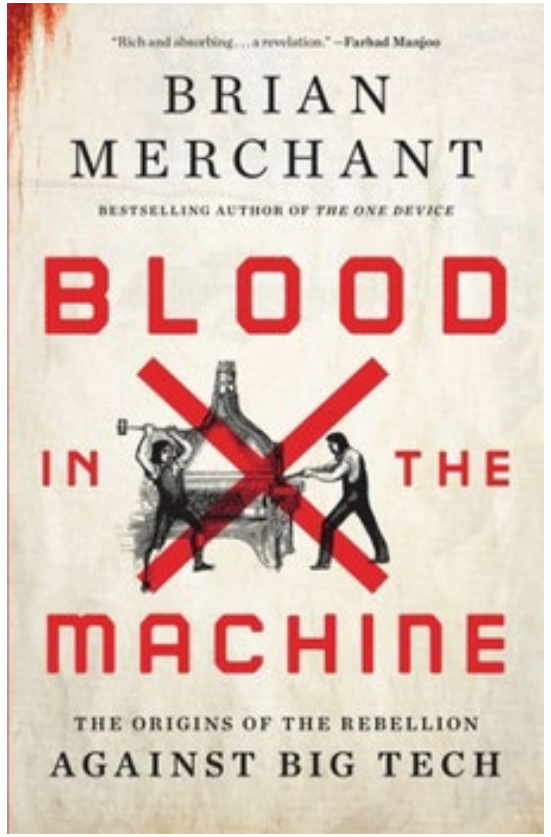
  "The law requires every judge to swear an oath to perform his or her work without fear or favor, but we must support judges by ensuring their safety. A judicial system cannot and should not live in fear."

https://www.supremecourt.gov/publicinfo/year-end/year-endreports.aspx

History: Luddites



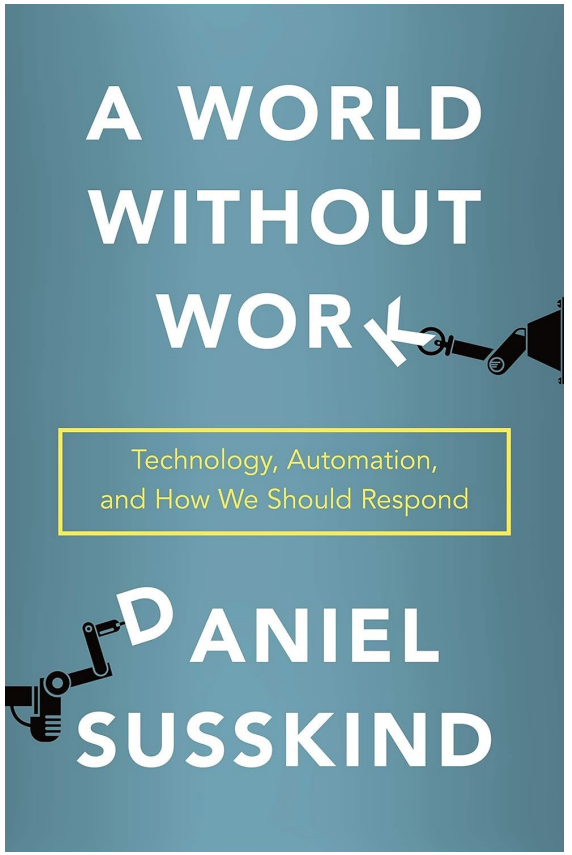
- Technology is good:
  - Luddite: Derogatory. A person opposed to new technology or ways of working*.
  - Automated weaving created new jobs: e.g., tailor
- Reality is not that simple:
  - Weavers were essentially part of the gentry
  - Opposed to changes in lifestyle, not technology
  - Factory work is **dehumanizing**
  - Factory quality was “good enough”
  - Transfer of wealth to factory owners



* Google dictionary, first definition.

Which jobs/tasks can be automated?

- The answer changes over time
- Today: any task that has
  - A clear evaluation of successful completion
  - Lots of training examples
- This will change



- Do you have a right to have a job?
- Are there other ways to have a rewarding life?

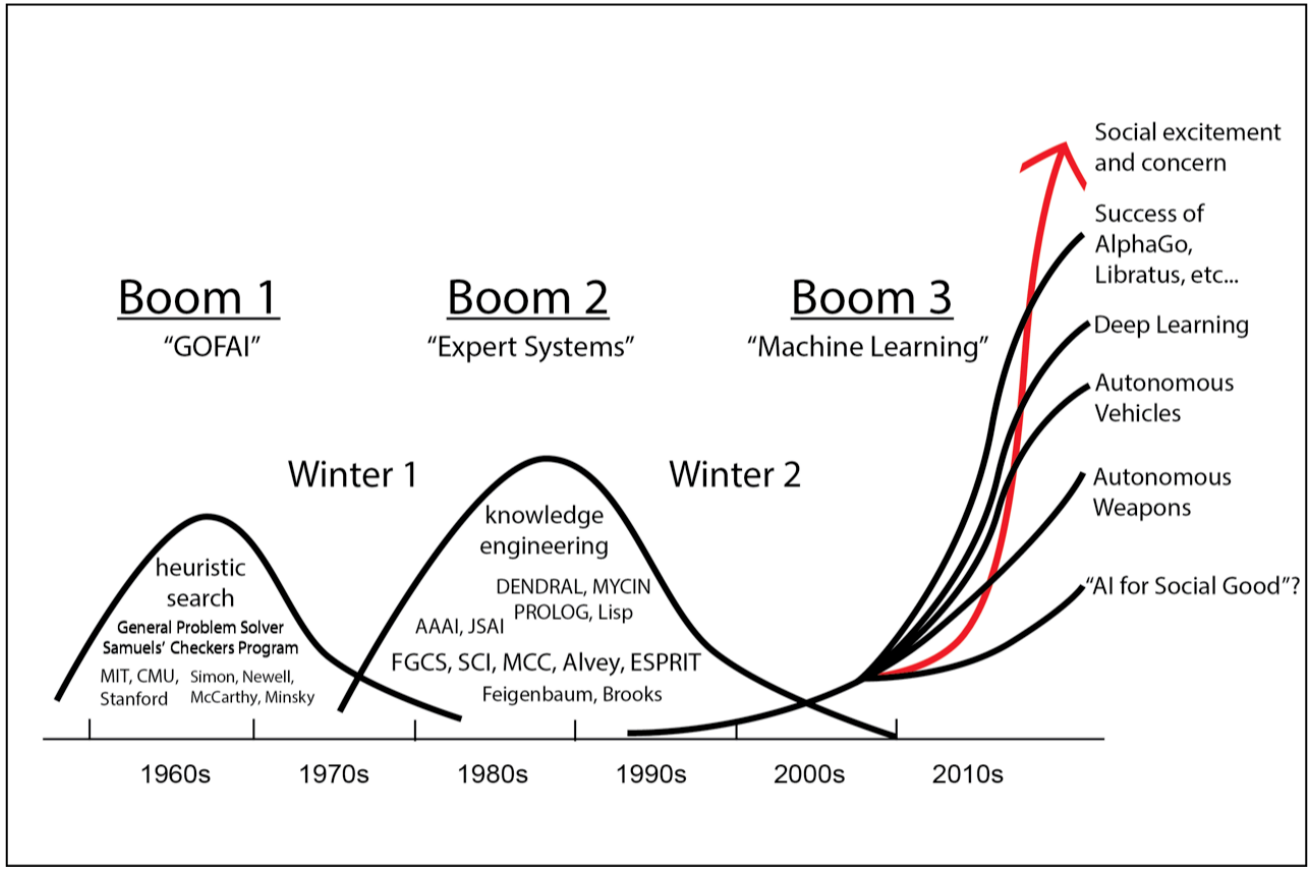
AI Overview

What is AI, and what makes this technology different?

Perspective: AI is Advanced Information technology

What is Artificial Intelligence?

- An aspiration, a grand challenge.
- How to measure success?
- The Turing Test – an imitation game.
- ML is learning statistical patterns
- **AI = Advanced Information** technologies



From Colin Garvey
(http://www.technologystories.org/ai-evolution/#_ftnref2)
GOFAI = Good Old Fashioned AI

Why is AI different?

- Scalability and ubiquity (minimal cost for replication across a wide range of platforms)
- Rapid evolution driven by advances in the underlying technologies, e.g., Moore's Law
- Algorithms, data, and information are not limited by physical "reality"
  - Shared learning (all autonomous vehicles learn from one another)
- Automating cognitive, not physical, labor

Advances in AI may outpace our ability to adapt societal norms

Concerns

Data

- Which communities are (not) represented in the data?
- Generating data is expensive; reusing existing data can be problematic.
- Privacy: potential to expose personal information.
- Sovereignty: ability to control how my data is being used
  - If you profit from using my data, am I entitled to a share?
  - What is the meaning of “fair use” in the use of data for training.

Equity

- High cost of entry means that models need to have commercial potential
  - Governments may subsidize costs for military or scientific goals
- Individual costs for using models
- Amplification of bias; self-fulfilling prophecies

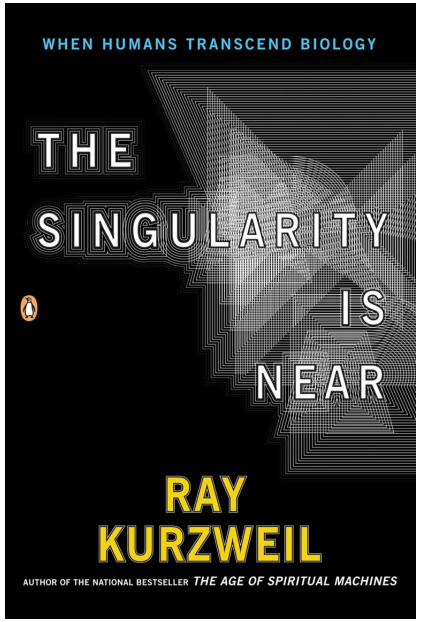
Explainability (telling stories)

- Learned models are based on identification of statistical patterns, e.g., predicting the next word in a sentence.
- There are no explanations, other than describing the statistical significance of the patterns.

- *F = ma* does not have an explanation (it's a law)
  - Simplicity
    - over fitting is unlikely
    - we can observe it over and over
  - Algebraic expression
    - we can carry it in our heads and
    - combine this model with others
  - Is there a "law" for Climate?

Superintelligence

- Artificial general intelligence
- Superintelligence: superhuman intelligence
  - What is the role of humans?
- Technological singularity – growth becomes uncontrollable and irreversible.



Powerful tools

- Information integrity: e.g., deepfakes and fake news.
- Enabling mass surveillance to identify and track individuals.
- (Cyber)security is becoming an arms race.
- Autonomous vehicles and weapons.

Executive and legislative activities

AI Bill of Rights; Executive Order 14110; AI Safety Institute

Legislative activity

Blueprint for an AI Bill of Rights (OSTP, October 2022)

- **Safe and Effective Systems**
  You should be protected from unsafe or ineffective systems
- **Algorithmic Discrimination Protections**
  You should not face discrimination by algorithms and systems should be used and designed in an equitable way
- **Data Privacy**
  You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used.
- **Notice and Explanation**
  You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you.
- **Human Alternatives, Consideration, and Fallback**
  You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter.

AI Bill of Rights – Commentary

Bouquets

- Important, positive, step forward
- Frames AI regulations as civil-rights issues, providing an important context for these regulations

Brickbats

- Should be connected to existing frameworks
  - Might be better understood as an interpretation of existing rights
- The nonbinding aspect makes this aspirational

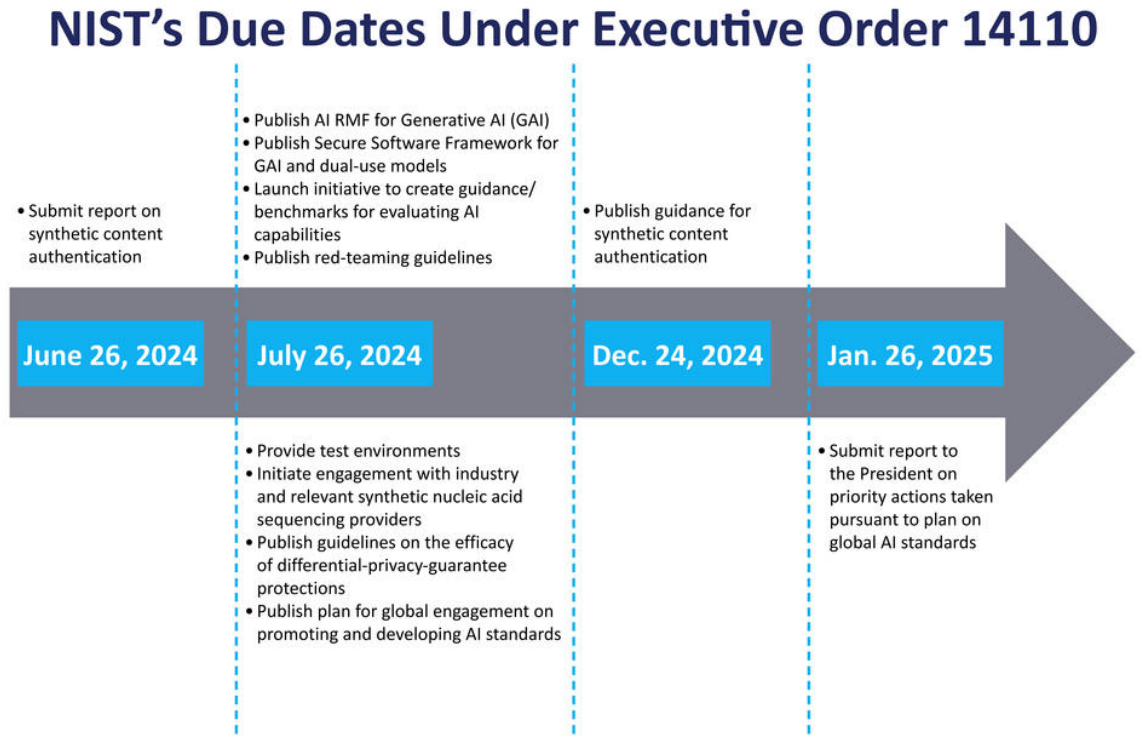
IEEE Spectrum, 14 Oct 2022, *6 Reactions to the White House's AI Bill of Rights: The nonbinding principles are being both celebrated and vilified.*

Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence (October 2023)

- Require that developers of the most powerful AI systems share their safety test results and other critical information with the U.S. government.
- Develop standards, tools, and tests to help ensure that AI systems are safe, secure, and trustworthy.
- Protect against the risks of using AI to engineer dangerous biological materials
- Protect Americans from AI-enabled fraud and deception by establishing standards and best practices for detecting AI-generated content and authenticating official content.

AI Safety Institute to be run by NIST

- Five Working groups
  - Risk Management for Generative AI
  - Synthetic Content
  - Capability Evaluations
  - Red-Teaming
  - Safety & Security
- AI Safety Institute Consortium
  - call is open until 15 Jan 2024



https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence

https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute

2024 is likely to see new laws regarding AI

- Artificial Intelligence (AI) Research, Innovation, and Accountability Act of 2023
  - **John Thune (R-SD)**, Amy Klobuchar (D-MN), Roger Wicker (R-MS), John Hickenlooper (D-CO), Shelley Moore Capito (R-WV), and Ben Ray Luján (D-NM)
  - Read twice and referred to the Committee on Commerce, Science, and Transportation
- Expect more action at the state level

Arizona HB 2685

- Second legislative session; summer 2022
- Original Intent: The legislature intends by this act to protect citizens of this state from potential harm caused by algorithmic artificial intelligence designed to emulate human beings.
- “Striker bill” – strike everything
  - Morphed into: Transportation tax; Maricopa county
  - Vetoed 2022-07-06

Roles for Society

The limited importance of expertise in technology.

Everyone

- Challenges and potential for societal impact do not require deep understanding of the technology.
- Critical concepts, like *fairness*, are social and not technical constructs.
- Do we need new rules or re-interpretation of existing rules?
- The potential for job displacement is real.

Enabling the conversation at UArizona

- Started in Summer of 2023: 60+ faculty, staff, and students
- Syllabus Guidance
  - polices need to be re-interpreted
  - instructors need to be clear about expectations
- Teams included: Industry, Equity, AI & Data Acumen, Training, Integrity, Syllabus Guidance, and Communications
  - https://artificialintelligence.arizona.edu
- Continued in Fall of 2023 with over 400 volunteers!

Questions?

Arthur “Barney” Maccabe
maccabe@arizona.edu

Coding Kindness: Regulatory Approaches to Technology and Compassion

EU GDPR (general data protection regulation)

- Everyone has the right to
  - the protection of personal data concerning him or her
  - access to data which has been collected concerning him or her, and the right to have it rectified
- This right is enshrined in article 8 of the Charter of Fundamental Rights.
  - Article 17: Right to erasure (“right to be forgotten”)
- Adopted 14 April 2016; became effective 25 May 2018
- Seems to have disproportionally adversely impact smaller companies*

Giorgio Presidente and Carl Benedikt Frey, “The GDPR effect: How data privacy regulation shaped firm performance globally,” VoxEU.org, 10 March 2022

Arizona HB 2685

Add Chapter 7 to the Arizona Revised Statutes

A. AN ALGORITHM THAT ENABLES ARTIFICIAL INTELLIGENCE TO LEARN AND IMPLEMENT DECISIONS WITHOUT HUMAN INTERVENTION MUST MEET THE FOLLOWING REQUIREMENTS:

1. SUPPORT HUMAN AGENCY AND FUNDAMENTAL RIGHTS.
2. COMPLY WITH ALL FEDERAL AND STATE LAWS.
3. FULFILL ETHICAL PRINCIPLES THAT ENSURE NO UNINTENDED HUMAN HARM OCCURS.
4. PROVIDE TRANSPARENCY AND TRACEABILITY OF DATA LOGS AND DECISION-MAKING.

B. ARTIFICIAL INTELLIGENCE MAY NOT INFRINGE ON A HUMAN BEING'S CONSTITUTIONAL RIGHTS.

The legislature intends by this act to protect citizens of this state from potential harm caused by algorithmic artificial intelligence designed to emulate human beings.

Generative AI and LLMs

Large-scale Language Models and ChatBots are one of the most visible AI technologies today. What are they and how are they built?

LLMs: Large-scale **Language** Models

- An LLM, e.g., GPT-4, is a statistical model of language
  - Trained to understand* human language: grammar, syntax, context, and nuances like sarcasm and humor
  - Based on predictive text generation
- A ChatBot, e.g., ChatGPT, is an application of an LLM
  - Additional training in conversation (context), may have safety features

Training, requiring large amounts of data and computation, has a high cost of entry

* Respond appropriately to an input

Extending LLMs

- Early LLMs required fine tuning for specific contexts.
  - Expensive in terms of data and computation
- Newer versions allow other methods
  - *Prompt engineering* is the creation of textual descriptions of tasks to be performed by the LLM
  - *Retrieval Augmented Generation* (RAG) uses retrieval from a knowledge store to augment predictive generation