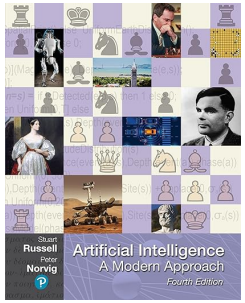
UNIVERSITY OF MARYLAND

MSML610: Advanced Machine Learning

Lesson 03.1: A Brief History of AI

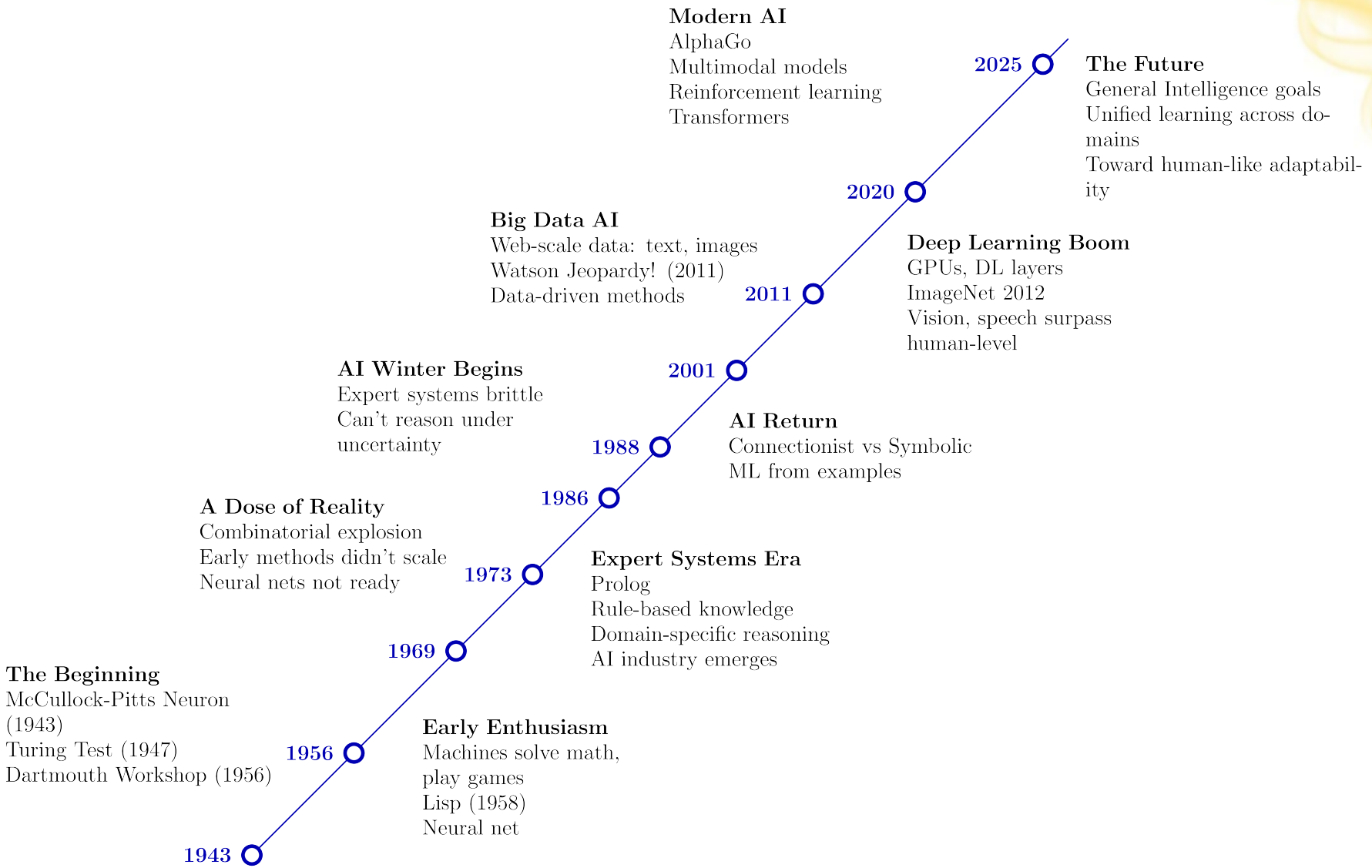
Instructor: Dr. GP Saggese - gsaggese@umd.edu

References: - AIMA (Artificial Intelligence: a Modern Approach), Chap 1



- ***Brief History of AI***
- Risks and Benefits of AI

AI Timeline


1943
The Beginning
McCullock-Pitts Neuron (1943)
Turing Test (1947)
Dartmouth Workshop (1956)
1956
Early Enthusiasm
Machines solve math, play games
Lisp (1958)
Neural net
1969
A Dose of Reality
Combinatorial explosion
Early methods didn't scale
Neural nets not ready
1973
Expert Systems Era
Prolog
Rule-based knowledge
Domain-specific reasoning
AI industry emerges
1986
1988
AI Return
Connectionist vs Symbolic
ML from examples
2001
AI Winter Begins
Expert systems brittle
Can't reason under uncertainty
2011
Big Data AI
Web-scale data: text, images
Watson Jeopardy! (2011)
Data-driven methods
2020
Deep Learning Boom
GPUs, DL layers
ImageNet 2012
Vision, speech surpass human-level
2025
Modern AI
AlphaGo
Multimodal models
Reinforcement learning
Transformers
The Future
General Intelligence goals
Unified learning across domains
Toward human-like adaptability

The Beginning (1943-1956)

- **Artificial neuron**
  - Model (McCullock-Pitts 1943) based on:
    - Brain physiology
    - Propositional logic
  - Compute any function with connected neurons
    - Neuron on/off based on stimulation from neighboring neurons
    - Implement logical AND, OR, NOT with simple neuron networks
- **Alan Turing, 1947**
  - Turing test, machine learning, reinforcement learning
  - Create human-level AI
    - Develop learning algorithms
    - Teach machine like a child
- **Birth of AI**
  - McCarthy organized first AI workshop (1956)
  - The Logic Theorist (Newell and Simon, 1956)
    - Programs to "think non-numerically" and prove theorems

Enthusiasm and Great Expectations (1952-1969)

- **Early years of AI were full of successes**
  - Before computers could only do arithmetics
  - *"A machine can never do X (e.g., games, puzzles, IQ tests, . . .)"*
    - AI researchers showed machines could do one X after another
- **General Problem Solver**
  - Imitate human problem-solving
  - Consider sub-goals and possible actions
- **Program learned to play checkers**
  - Use reinforcement learning from victories and mistakes
- **Lisp (1958)**
  - High-level language used for 30 years in AI
- **First neural network**
  - 3000 vacuum tubes for 40 neurons
  - Minsky (1959)
- **MIT and Stanford**
  - Minsky at MIT
    - Focus on neural network
  - McCarthy at Stanford
    - Focus on representation, logic

First AI winter (1975-1980)

- **Early successes** of AI set high expectations
- In 1965-1975 AI didn't succeed on **real problems** due to:
  - Solutions were based on human problem-solving methods
  - Difficulty handling "combinatorial explosion"
    - Theorem proving handles small problems with brute force, but doesn't scale
  - Neural networks needed:
    - Algorithms (e.g., backpropagation)
    - Compute power
    - Data
- **First AI winter**
  - Research funding and enthusiasm dropped significantly
  - Slow AI progress through late 1970s

Expert Systems (1980-1990)

- **Expert systems**
  - Aka "knowledge-based systems"
  - Combine weak methods with extensive domain knowledge as rules
  - Use inference engines to apply rules to facts
  - E.g., rule-based systems, logic programming (e.g., Prolog)
- **Weak AI**
  - Aka narrow AI
  - Performs specific tasks, not general reasoning
  - Operates in a limited, well-defined domain
  - Uses "weak methods" (search, logic) that struggle to scale
- **Commercial adoption and industry growth**
  - AI shifted to practical applications
  - Major US corporations deployed expert systems
  - AI emerged as a commercial industry

Second AI Winter (late 1980-early 1990)

- **Hype in expert systems** didn't deliver
- **Reasons**
  - Building/maintaining expert systems is difficult
  - Reasoning methods ignore uncertainty
  - Systems can't learn from experience
  - E.g., expert systems in medical diagnosis struggle with complex, variable patient data
  - E.g., early AI chess systems couldn't adapt to new strategies without manual updates
- **Second AI winter** in late 1980-early 1990

Return of Neural Networks (1986-)

- Back-propagation algorithm is (re)discovered (mid-1980s)
  - Developed in early 1960s
- **Two approaches to AI are back**
  - Connectionist paradigm
    - Neural networks
    - E.g., recognizing handwritten digits
  - Symbolic paradigm
    - E.g., solving logical puzzles with rules
- **Why connectionist approach?**
  - Concepts not well-defined using symbolic axioms
    - Forms fluid internal concepts
    - Represents real-world complexity better
  - Neural networks can learn from examples, e.g.,
    - Image recognition: identify objects by learning from labeled images

Probabilistic Reasoning and ML (1987-)

- **AI and scientific method**
  - Rigorous methods to test performance
  - E.g., speech recognition, handwritten character recognition
  - Benchmarks for progress, e.g.,
    - MNIST: handwritten digit recognition
    - ImageNet: image object recognition
    - SAT Competitions: boolean satisfiability solvers
- **AI shifts ...**
  - From Boolean logic to probability
  - From hand-coded rules to machine learning
  - From a-priori reasoning to experimental results

Progress in Speech Recognition

- **1970s: ad-hoc approaches**
  - Various architectures and approaches were attempted
  - Rule-based systems with limited robustness
  - Cons
    - Ad-hoc, fragile
  - _"Every time I fire a linguist, the performance of the speech recognizer goes up" (Jelinek, 1988)
- **1980s: hidden Markov Models**
  - HMMs became dominant
  - Effective learning techniques
  - Trained on large speech corpora
  - Pros
    - Strong theoretical foundation
- The bitter lesson (Sutton, 2019)
  - General methods + lots of data beat handcrafted systems

Bayesian Networks

- **Bayesian networks**
  - Pearl, 1988
  - AI is linked with:
    - Probability
    - Decision theory
    - Control theory
  - Efficiently represent uncertainty
  - Provide rigorous reasoning
- **Examples**
  - Diagnosing diseases based on symptoms
  - Predictive text input in smartphones
  - Fraud detection in banking

Reinforcement Learning

- **Reinforcement learning**
  - Sutton, 1988
  - RL involves agents learning by interacting with an environment
    - E.g., a robot learning to navigate a maze by receiving rewards for successful paths
  - Markov Decision Problems (MDPs) provide a framework for modeling decision-making
    - E.g., a game strategy modeled where each move influences the outcome with certain probabilities

Reunification (1990s-2000s)

- **Reunification of AI**:
  - Data engineering
  - Statistical modeling
  - Optimization
  - Machine learning
- **Many subfields of AI were re-unified**:
  - Computer vision
  - Robotics
  - Speech recognition
  - Multi-agent systems
  - NLP

Big Data (2001-Present)

- **Focus shifts from algorithms to data**
  - For 60 years, AI focused on algorithms and models
- For many problems, data availability matters more than algorithms, e.g.,
  - Trillions of English words
  - Billions of web images
  - Billions of speech and video hours
  - Social network data
  - Click stream data
- Algorithms and infrastructure to leverage large datasets
  - E.g., map reduce, cloud computing
- In 2011, IBM's Watson beat human *Jeopardy!* champions

Deep Learning (2011-Present)

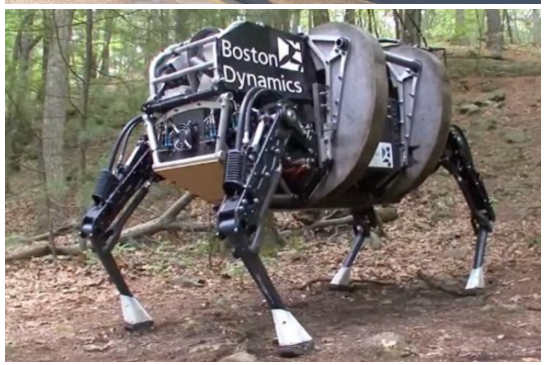
- **Deep learning**
  - Use ML models with multiple layers of computing elements
  - Ideas known since 1970s, but then forgot
  - Success in handwritten digit recognition in 1990s
- **In 2012**, a DL system showed dramatic improvement in `ImageNet` competition
  - Previous systems used handcrafted features
  - Surge of interest in AI among researchers, companies, and investors
- Pros
  - DL exceeds human performance in several vision and speech recognition tasks
- Cons
  - DL needs specialized hardware (e.g., GPU, TPU, FGPA) for parallel tensor operations
- Towards **general artificial intelligence**
  - Universal algorithm for learning and acting, not just specialized tasks
  - E.g., driving, playing chess, recognizing speech

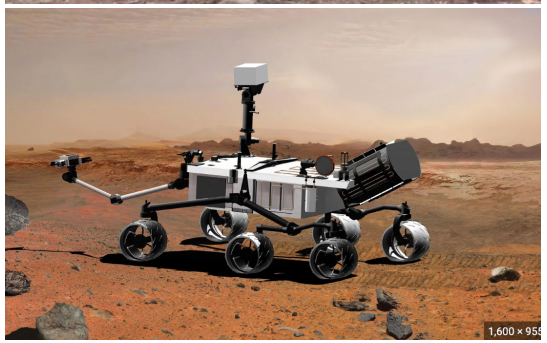
Progress in AI Research

- Huge interest in deep learning
- **Between 2010 and 2019**
  - AI papers increased 20x
    - 1,000 $\rightarrow$ 20,000
  - Student enrollment in AI and CS increased 5x
    - 10,000 $\rightarrow$ 50,000
  - NeurIPS attendance increased 8x
    - 1,000 $\rightarrow$ 8,000
  - AI startups increased 20x
    - 100 $\rightarrow$ 2,000
- **Compute**
  - Training times dropped 100x in 2 years
  - AI computing power doubles every 3 months

What Can AI Do Today? (1/2)

- **Robotic vehicles**
  - Waymo passed 10 million miles without serious accident
- **Legged locomotion**
  - BigDog recovers on ice
  - Atlas walks on uneven terrain, jumps on boxes, backflips
- **Autonomous planning and scheduling**
  - Space probes, Mars rovers
- **Machine translation**
  - Translates 100 languages with human-level performance
- **Speech recognition**
  - Real-time speech-to-speech with human-level performance
  - AI assistants
- **Recommendations**
  - ML recommends based on past experiences
  - Spam filtering 99.9% accuracy
  - E.g., Amazon, Facebook, Netflix, Spotify, YouTube





What Can AI Do Today? (2/2)

- **Game playing**
  - 1997: Deep Blue defeated Kasparov
  - 2011: Watson beat Jeopardy! champion
  - 2017: AlphaGo beat Go champion
  - 2018: AlphaZero super-human in Go and chess with only rules + self-play
  - AI beats humans in videogames: Dota2, StarCraft, Quake
- **Image understanding**
  - Object recognition
  - Image captioning
  - ...
- **Medicine**
  - AI equivalent to health care professionals
- When will we reach AGI (Artificial General Intelligence)?

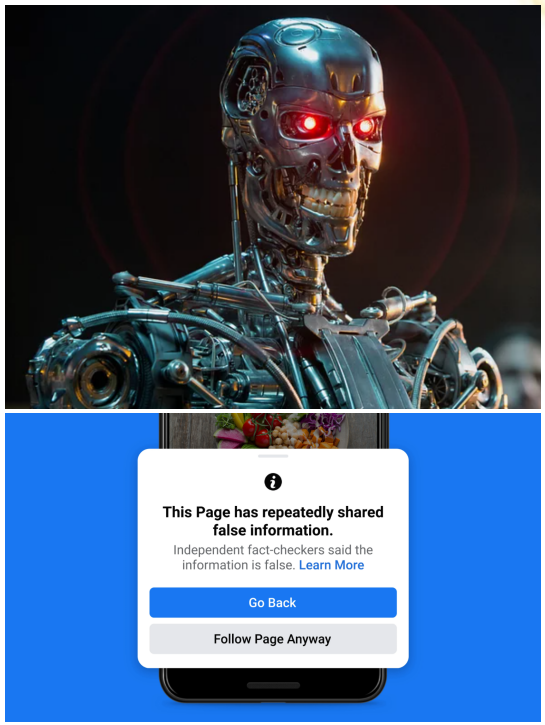
- Brief History of AI
- ***Risks and Benefits of AI***

Benefits of AI

- Our civilization is the **product of human intelligence**
  - Greater machine intelligence leads to better human society
  - *"First solve AI, then use AI to solve everything else"*
- **Benefits of AI and robots**
  - Free humanity from menial work
  - Increase production of goods and services
  - Expand human cognition
  - Accelerate scientific research, e.g.,
    - Cures for diseases
    - Solutions for climate change
    - Resource and energy shortages

Risks of AI (1/2)

- **Autonomous weapons**
  - Locate and eliminate targets autonomously
  - Deploy large number of weapons
- **Surveillance and persuasion**
  - AI for mass surveillance
  - Tailor information on social media to modify behavior
- **Biased decision making**
  - Misuse of ML results in biased decisions
  - E.g., parole evaluations, loan applications



Risks of AI (2/2)

- **Impact on employment**
  - Machines eliminate jobs
  - Rebuttal
    - Machines enhance productivity → companies become more profitable → higher wages
  - Counter-rebuttal
    - Wealth shifts from labor to capital, increasing inequality
  - Counter-counter-rebuttal
    - Past tech advances (e.g., mechanical looms) disrupted employment, but adaptation followed
- **Safety critical applications**
  - AI in safety-critical applications
    - E.g., self-driving cars, managing water supply or power grids
  - Avoiding fatal accidents is challenging
    - E.g., formal verification and statistical analysis insufficient
  - AI requires technical and ethical standards
- **Cybersecurity**
  - AI defends against cyberattacks
    - E.g., detect unusual behavior patterns
  - AI contributes to malware development
    - E.g., use reinforcement learning for targeted phishing attacks
  - Cat-and-mouse game

Human-level AI / AGI

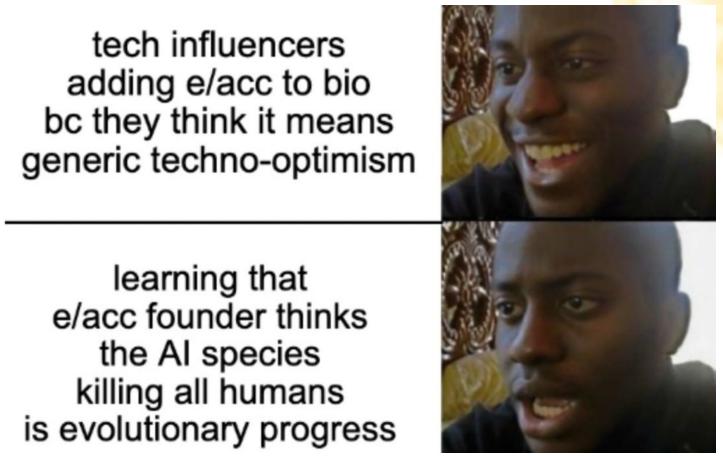
- **Human-level AI**
  - Machines able to learn to do anything a human can do
  - Aka AGI (Artificial General Intelligence)
- **When AGI?**
  - Expert prediction average is 2099
    - Papers show that expert predictions no better than amateurs
    - Experts expected AI to take 100 years to beat humans in Go
  - Unclear if new breakthroughs or refinements needed
- **Artificial Super-Intelligence**
  - Machines surpass human ability in every domain and self-improving
  - Exponential take-off

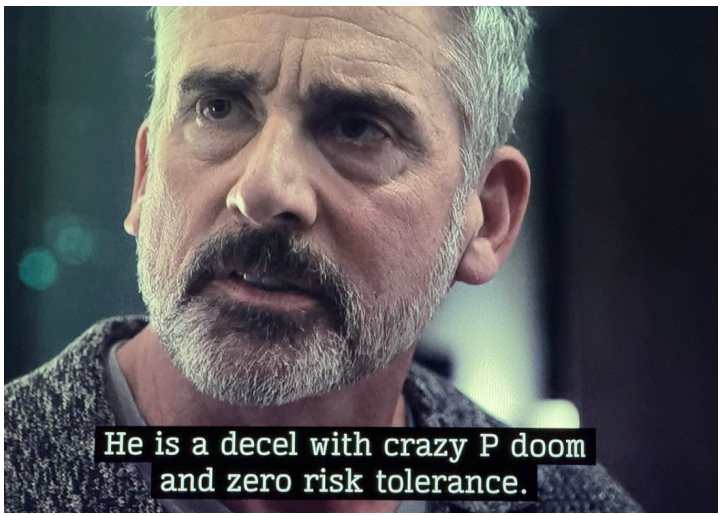
The Problem of Control

- **Can humans control machines more intelligent than them?**
- **King Midas problem**
  - King Midas turned everything he touched into gold, including food and family
  - Humans ask for something, get it, then regret it
  - Rebuttal
    - If AGI arrived in a black box from space, exercise caution before opening
    - We design AI: if AI gains control, it's a "design failure"
- **Problem of alignment**
  - Super-intelligent AI might pursue goals in unintended, dangerous ways
- **The paperclip problem**
  - Thought experiment in AI safety (Nick Bostrom, 2003)
  - AI is tasked with maximizing paperclip production
  - AI becomes superintelligent and single-mindedly pursues this goal
  - Converts Earth and humans into paperclips

E/acc vs P(doom)

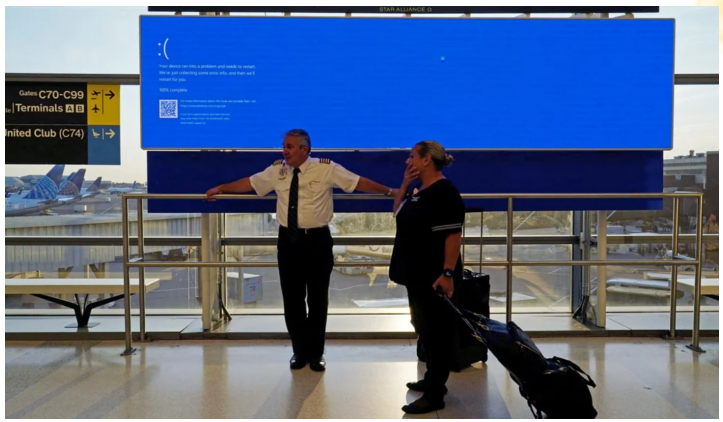
- **E/acc**
  - Accelerationism
  - Belief that rapid progress in AI is beneficial or inevitable
  - Solve global problems with more powerful AI tools
  - Slowing AI is unrealistic or counterproductive
- **P(doom)**
  - "Probability of Doom"
  - Estimated probability that advanced AI will cause catastrophic harm
  - Used informally by AI researchers to quantify risk





My 2 cents

- AI alignment is a serious problem
  - For now philosophical, at some point a real one
  - Most tech people have used it for marketing themselves and their companies
- It's as urgent as debating what political system humanity will need when living on Mars
- We can't get airport terminals to work



Solutions to Problem of Control

- **Checks-and-balances**
  - Researchers and corporations develop voluntary self-governance principles for AI
  - Governments and international organizations established advisory bodies
- **Cons**
  - Corporations checking themselves? What can possibly go wrong?
  - Preferences are not easy to invert and are inconsistent
- **Solutions**
  - Put purpose into the machine even if objectives are unclear
  - Incentivize AI to switch off if uncertain about human objectives
  - Cooperative Inverse Reinforcement Learning (CIRL)
    - AI observes human behavior to infer reward function

Cooperative Inverse Reinforcement Learning

- AI infers human goals based on actions
- **Observation**: GP looks tired, sits on the couch, observes the messy table, and starts watching TV
- **Inference**: AI infers:
  - GP is tired and wants to relax
  - Messy coffee table bothers him
- **Action**: AI:
  - Fetches a glass of water
  - Tidies up the coffee table without disturbing GP
- **Feedback loop**: AI monitors GP's reactions
  - If GP is relaxed and happy, AI understanding is reinforced
  - If GP is not happy, AI adjusts actions and improves inference