RSAConference2022

San Francisco & Digital | June 6 – 9

TRANSFORM

SESSION ID: **MLAI-T08**

AI: Legal and Policy Considerations and Landmines to Avoid

Behnam Dayanim

Global Chair, Data Privacy & Cybersecurity Practice, and Chair, Advertising & Gaming Practice
Paul Hastings LLP

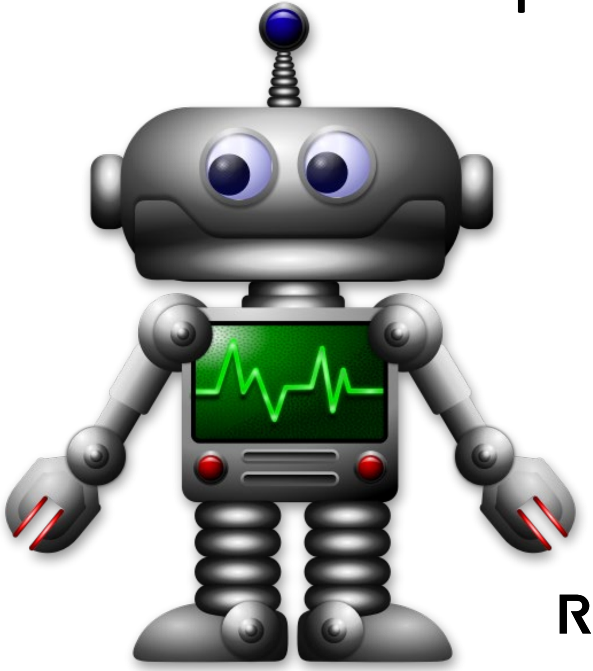
Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

What to Expect?

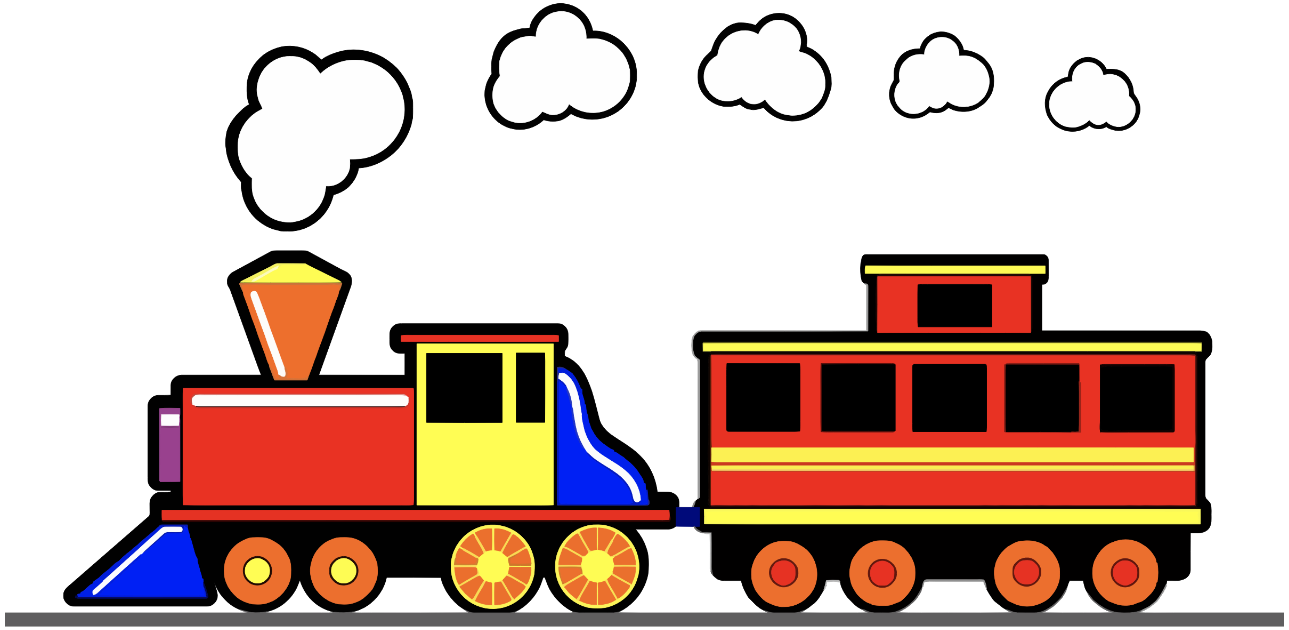
- What is AI, and what should we expect from it?
- How should we think about choices and trade-offs?
- Congress & the States
- The new EU Artificial Intelligence Act and its implications for AI globally
- What to do?



What is AI?

AI isn't a thing, like a train, but rather a set of techniques aimed at approximating some aspect of cognition…..

Ryan Calo – UW School of Law



What is AI?

- Any artificial systems that perform tasks under varying and unpredictable circumstances, without significant human oversight, or that can learn from their experience and improve their performance. . . . They may solve tasks requiring human-like perception, cognition, planning, learning, communication, or physical action. **In general, the more human-like the system within the context of its tasks, the more it can be said to use artificial intelligence**.
- Systems **that think like humans**, such as cognitive architectures and neural networks.
- Systems **that act like humans**, such as systems that can pass the Turing test or other comparable test via natural language processing, knowledge representation, automated reasoning, and learning.
- A set of techniques, including machine learning, **that seek to approximate some cognitive task**.
- Systems **that act rationally**, such as intelligent software agents and embodied robots that achieve goals via perception, planning, reasoning, learning, communicating, decisionmaking, and acting.
  - *FUTURE of Artificial Intelligence Act of 2020 (not passed)*

What is AI?

“**General**” versus “**narrow**”

- General – a notional future artificial intelligence system that exhibits apparently intelligent behavior **at least as advanced as a person** across the range of cognitive, emotional, and social behaviors.
- Narrow - an artificial intelligence system that addresses **specific application areas** such as playing strategic games, language translation, self-driving vehicles, and facial or other image recognition

What is AI?

- AUTOMATED DECISION SYSTEM.—The term “automated decision system” means any system, software, or process (including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques and excluding passive computing infrastructure) that uses computation, the result of which serves as a basis for a decision or judgment.
  - *Algorithmic Accountability Act of 2022 (introduced)*

What is AI?

- A system that (i) receives machine and/or human-based data and inputs, (ii) infers how to achieve a given set of human-defined objectives using learning, reasoning or modelling implemented with the techniques and approaches listed in Annex I, and (iii) generates outputs in the form of content (generative AI systems), predictions, recommendations or decisions, which influence the environments it interacts with
  - Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
  - Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
  - Statistical approaches, Bayesian estimation, search and optimization methods.
- **Excludes from scope**: military/national security, pure scientific research/development or any R&D so long as AI system not placed on market or into service
- *EU proposed Artificial Intelligence Act (Nov. 2021 “Presidency Compromise Draft”*)

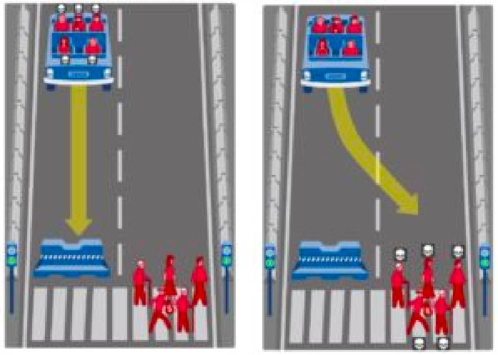
Is that really what we want?

- Human=good?
- Or do we demand better than human?
- “Backseat Driver” Phenomenon
- *Bias is neither new nor unique to AI and it is not possible to achieve zero risk of bias in an AI system.* (NIST, March 2022)

The Dilemma

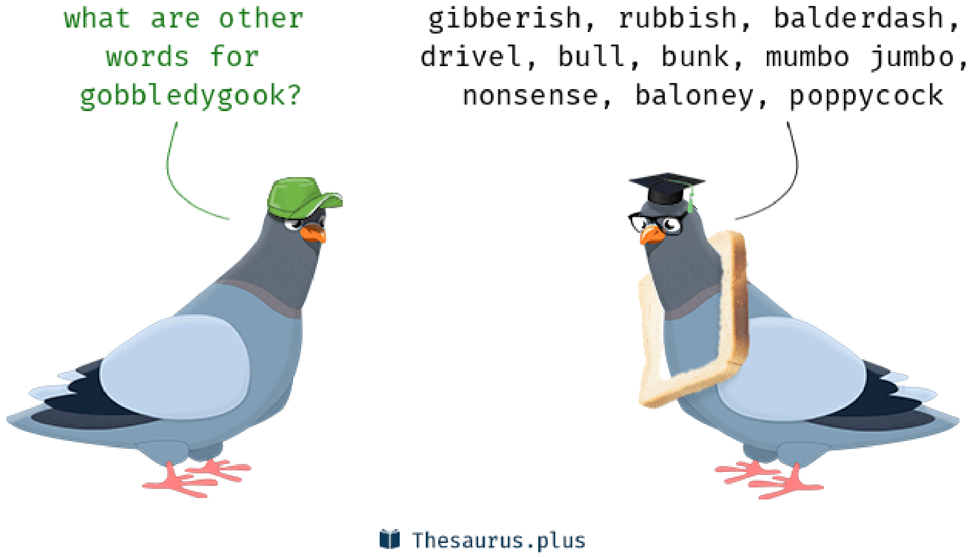
- Autonomous vehicles
- What does safety mean? For whom and under what circumstances?
- Moralmachine.net
- Is crowdsourcing the answer (i.e., kill the elderly)?

What should the self-driving car do?



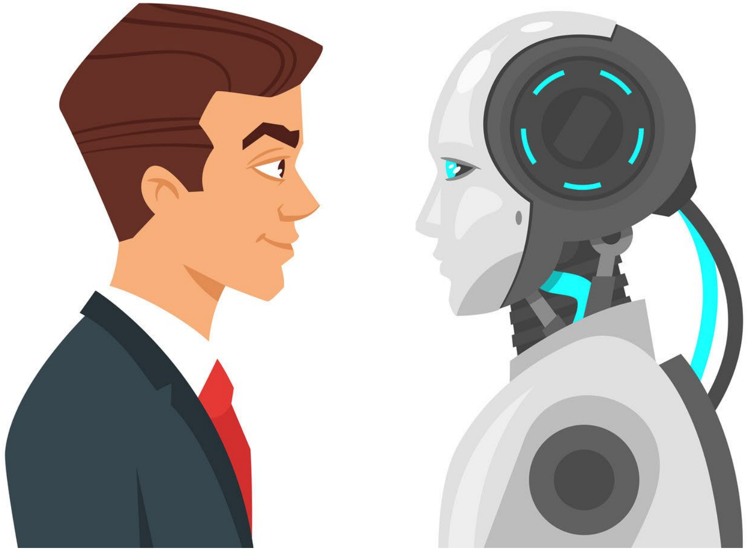
It’s not just cars….

- Workplace performance monitoring
- Video-chat emotional AI
- Language biases in content moderation and monitoring

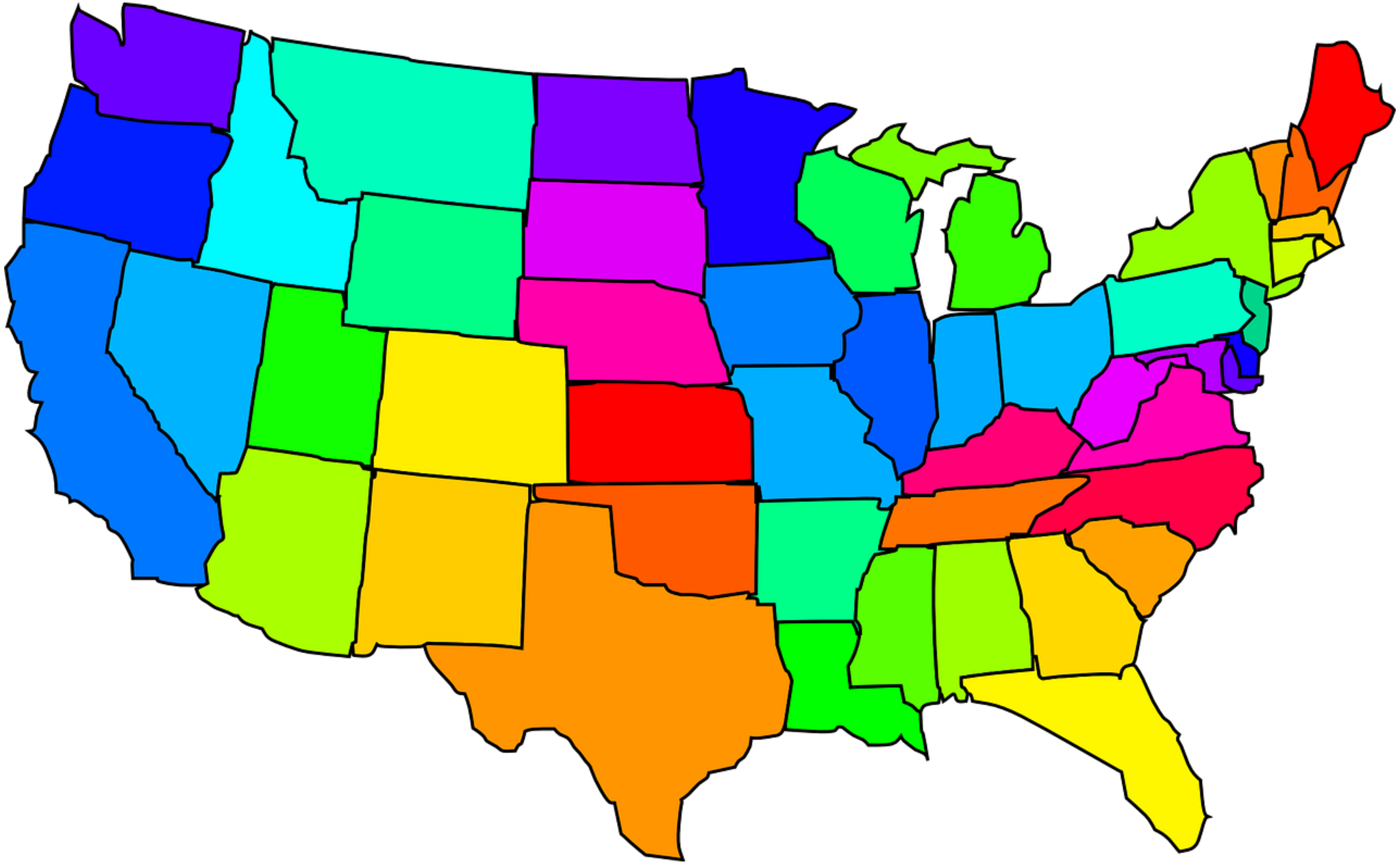

what are other words for gobbledygook?
gibberish, rubbish, balderdash, drivel, bull, bunk, mumbo jumbo, nonsense, baloney, poppycock
Thesaurus.plus

What Works Best?

- *Kartik Hosanagar – Wharton School of Business, UPenn*
- Research on AI/human interactions
- AIs know when to delegate
- Humans don't



Legal Framework - US



Recent Actions

- Feb. 2019 - E.O. 13859; established the American AI Initiative
  - Focused on increasing investment in AI
  - Brief mention of privacy, civil liberty and safety/security but no direction
- OMB guidance in late 2020 warned against stifling AI adoption, emphasized non-regulatory approaches to address AI risk and articulated 10 "stewardship" principles

- Public trust in AI
- Public participation
- Scientific integrity and information quality
- Risk assessment and management
- Benefits and costs
- Flexibility in development – technology neutrality
- Fairness and non-discrimination
- Disclosure and transparency
- Safety and security
- Interagency coordination

Executive Action

- Most active agency – Federal Trade Commission
  - FTC Act
  - Fair Credit Reporting Act
  - Equal Credit Opportunity Act
- Articulated several core principles:
  - *Start with the right foundation*
  - *Watch for discriminatory outcomes*
  - *Embrace transparency and independence*
  - *Don’t exaggerate*
  - *Tell the truth about how you use data*
  - *Do more good than harm*
  - *Hold yourself accountable*

Executive Action

- FTC Everalbum settlement, Jan. 2021
- Photo app developer – app “Ever”
- FTC alleged that misrepresented it would not use facial recognition unless users opted in
  - Except IL, WA, TX and EU
- Settlement required deletion of all photos obtained without consent, all “Face Embeddings” and “Affected Work Product”
  - Data derived from an image of a face
  - Models or algorithms developed in whole or in part using the images

Executive Action

- WW International – FTC March 2022
- Kurbo healthy eating app targeted toward kids
- $1.5M penalty
- Algorithmic destruction
- Important new tool for agency

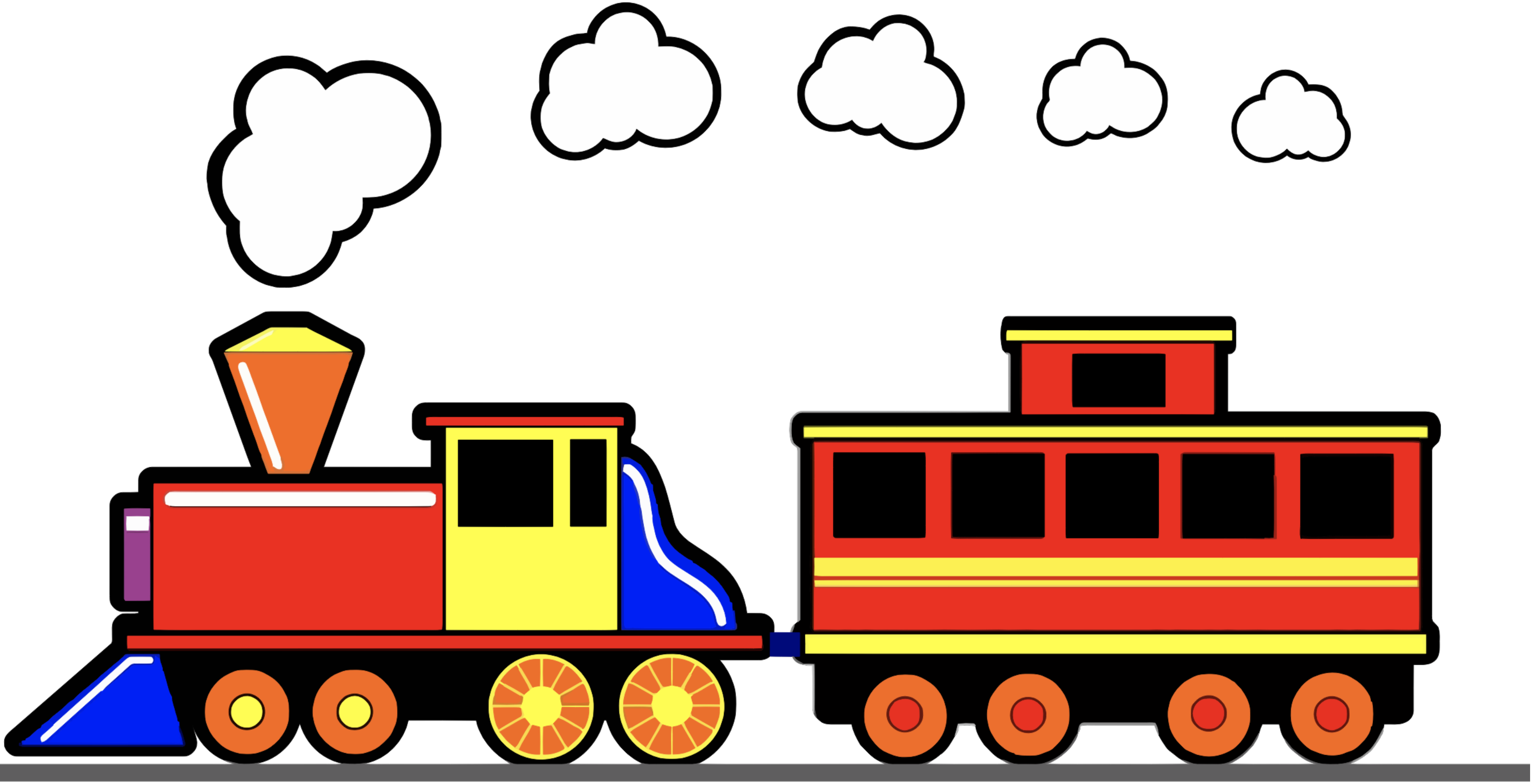
Not Just the FTC Any Longer…

- CFPB – March 2022 update to examination manual to include routine review of use of models/algorithms, input data collected, policies to review decision-making for discriminatory impact
  - Also outlined proposal in Feb 2022 to prevent bias in home evaluation algorithms ("automated valuation models"), in conjunction with March 2022 interagency plan to address property and valuation equity (13 agencies)
  - Invoke UDAAP authority
- October 2021, White House Office of Science and Technology Policy announced intent to develop "AI Bill of Rights"
- March 2021, federal financial regulators sought comment on governance and controls over financial institutions' use of AI

NIST Contribution

- March 2022 report identifies three key sources of bias:
  - Systemic – institutional, historical, societal
  - Computational and statistical – data sets/samples that are not representative
  - Human – individual and group, generally implicit
- Danger of “techno-solutionism”
- Quantitative approaches are not always best
- Must evaluate, document and measure “real-world value” of an AI system
- Forming a voluntary risk management framework for AI

So… Is it a Train?



Legislation lags . . . for now

- Algorithmic Accountability Act of 2022
  - Requires companies to assess impacts of automated critical decision-making
    - Companies subject to FTC and with more than $50M annual revenue or $250M equity value, or PI of more than 1M consumers, households or devices for purpose of developing or using AI
    - “Critical” means “legal, material or similarly significant effect” on a consumer’s life relating to access to or cost, terms or availability of education, employment, essential utilities, family planning, financial services, healthcare, housing, legal services or other comparable services
  - FTC to write rules for assessment and reporting, and to stand up a 50-person Bureau of Technology to enforce the act

Other federal bills

- Several bills to fund research, including to increase cooperation among allies, to counter China, etc.
  - One bill would enable AI to prescribe drugs if authorized by State and FDA
- Artificial Intelligence for Agency Impact Act (introduced)
  - Require each agency to assess adoption of AI, including ethical issues presented
  - Similar bill also introduced by Rep Maloney (D-NY) to strengthen Privacy & Civil Liberties Oversight Board's oversight of counterintel use of AI
- AI Training Act – passed Senate
  - Directs OMB to develop and all agencies (other than DoD and National Nuclear Security Admin) to implement training, including of risks to privacy and discrimination

State activity

- Lots of bills, but few have passed
- Existing law – Illinois AI Video Interview Act
  - Requires consent to use AI in job applicant interviews
  - If rely solely on AI, must collect and report on race/ethnicity
- Colorado SB 21-169, enacted last June: bans insurers from using algorithms and predictive models that use external consumer data if result is to unfairly discriminate based on a protected category

State Initiatives

- Other states considering action include California, New Jersey, Washington

The European Union

European Union AI Act (proposed)

- *Laying Down Harmonized Rules on Artificial Intelligence*
- Prohibited Uses
- High-Risk AI Systems
- Limited Risk Systems
- Minimal Risk Systems
- European Artificial Intelligence Board and national supervisory authorities
- Fines of up to 6% global revenue

Prohibited Uses

- Subliminal techniques to distort behavior
- Targeting human vulnerabilities
- Social scoring
- Real-time remote biometrics in public spaces for law enforcement *unless strictly necessary*

High-Risk Systems

- Very broad, including

  Biometrics

  Critical infrastructure

  Educational access

  Work-related decisions

  Public assistance

  Law enforcement

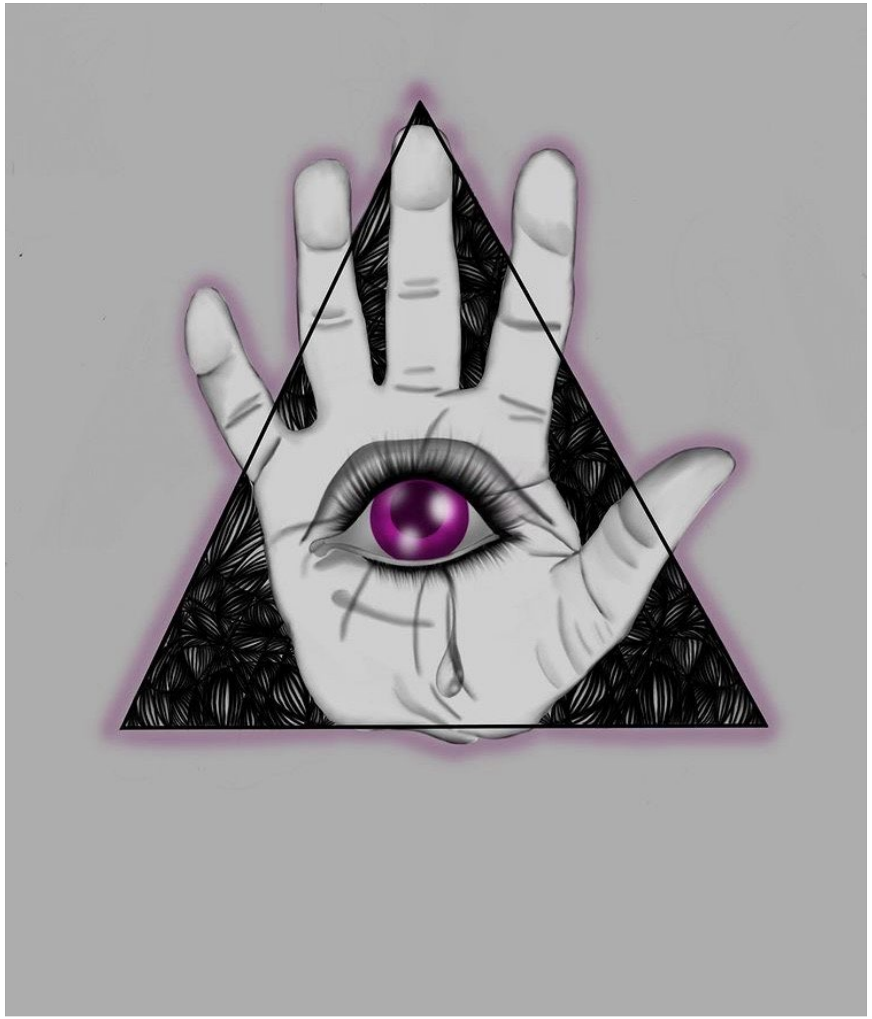
  Border control

  Judicial settings

High-Risk Systems

- Ongoing risk management
- Data quality and governance
- Record-keeping and transparency
- Human oversight
- Accuracy, robustness and cybersecurity
- All systems must be registered

So Now What?



Insta: Artby_dd

Four pillars of artificial intelligence governance

- **Fairness (Effectiveness)**
  - Implementing procedures and standards to ensure systems are meeting their intended purpose without bias.
- **Competence**
  - Specifying the knowledge and skills required of human developers and operators of AI to promote trust in AI’s use and mitigate the risk of bad outcomes.

Four pillars of artificial intelligence governance

- **Transparency**
  - Revealing appropriate information about systems to develop trust in and respect for systems and their capabilities, including through appropriate oversight.
- **Accountability**
  - Understanding who is responsible for the implementation and outcomes of systems, and creating the ability to apportion responsibility for outcomes. Requires measurability.

Four pillars of artificial intelligence governance

- Not all AI applications require equally strong or similarly structured governance programs.
- Strength in one pillar may reduce need for strength in others.

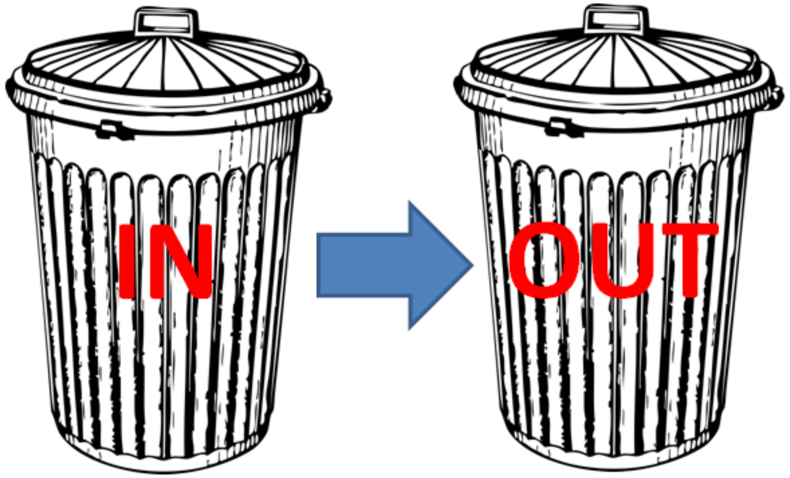
Designing governance mechanisms: Fairness/Effectiveness

- Implement procedures to measure effectiveness of the system, such as single-system validation or benchmarking exercises.
  - Single-system: measures effectives of one app through sampling and review of the data the system processed and the outcomes it reached.
  - Multi-system (benchmarking): comparative evaluation of systems meant to serve the same business objective.
- When defining metrics:
  - Keep in mind your requirements and evaluate what data are available
  - Seek input from stakeholders
  - Use metrics acceptable to relevant commercial or research communities

Designing governance mechanisms: Fairness/Effectiveness

- Drive robust human-led sampling and auditing, including working with curated data sets

- Bigger data may not be better data
  - Quality matters; garbage in/garbage out
  - Paralysis by analysis



Designing governance mechanisms: Effectiveness

- Imperative to Combat Bias in Algorithmic Decision-Making
- Ethical imperative
- Legal imperative
- Carefully consider the diversity of and hidden biases in your dataset
- Examples:
  - Using past medical costs to predict the need for future health interventions
  - Distance from employment site as factor in job consideration
- Audit algorithmic outcomes on key subpopulations with external datasets

Designing governance mechanisms: Competence

- Think pilots, neurosurgeons, nuclear launch operators….
- Competence standards are necessary to have effectiveness and trust, and help contain liability
- Tailored to the application and potential risk
- Training and skills required for system developers and operators
- Avoid groupthink – diversity of the team is critical

Designing governance mechanisms: Competence

- Safeguards against improper use or operation of the system
- Defined circumstances when system operator should override the system
- AI and Us – a **team**; human-centric design and oversight

- Continuous knowledge development
  - TEVV
  - Procedures for gauging the effectiveness of the system
  - Considerations to take into account when interpreting the results

Designing governance mechanisms : Transparency

- Transparency builds trust, increases accountability, and creates opportunities for course correction
  - Consider what categories of information about the system and its performance can be disclosed
  - IP can and should be protected
- Contractual provisions and whistleblower protections should be addressed
- Consider human-led appellate processes for algorithmic decisions, including independence and publication guidelines (remember, the "**TEAM**")
- Transparency reports
- But not a cure-all: an incomplete window limited by the AI objectives

Designing governance mechanisms : Accountability

- Clearly define success – what is a “right answer”?
- Well-documented lines of responsibility among those who develop and operate the system
- Training to ensure that those engaged in the adoption and operation of the system understand their responsibilities and the potential liability that could arise out of the system’s results
- Contractual provisions specifying responsibility for functions and outcomes
- Internal oversight mechanisms, audits and verification

What to Do? (The Apply Slides!)

"AI's ability to spot patterns makes it increasingly possible to derive the intimate from the available." (Ryan Calo)

- Data Protection offers a useful framework for thinking about AI
- "**AI impact assessment**"

What to Do? (The "Apply" Slides!)

- Focus on the narrow – clearly define the objective
- Consider how to avoid mission creep
- Purpose – well-defined?
- Competence and inclusivity – capable and diverse inputs?
- Can you **articulate** clearly and plainly how it works?
- Is there a "failsafe"? Human appeal?