Limiting or Refining the Digitization of Judgment: Two Paths for Regulating AI in the Workplace
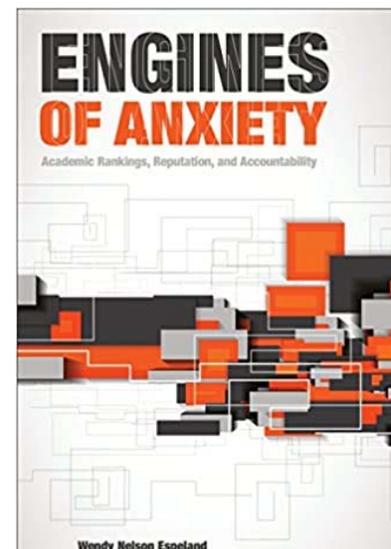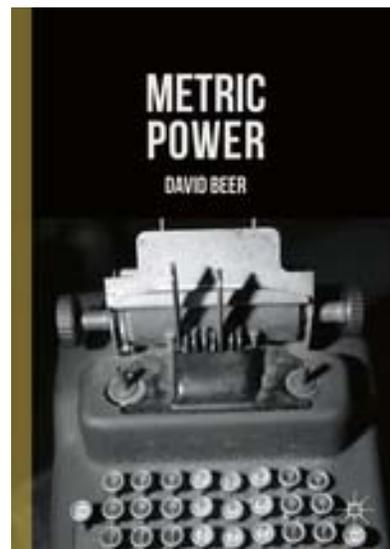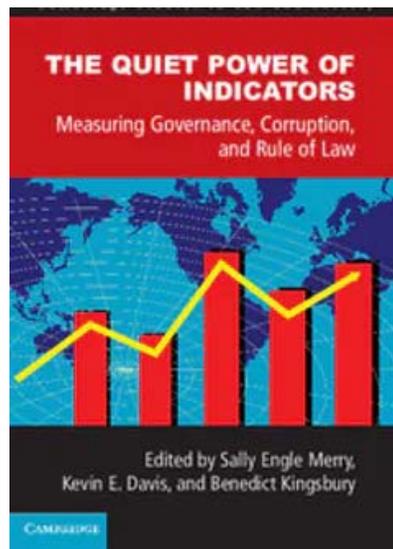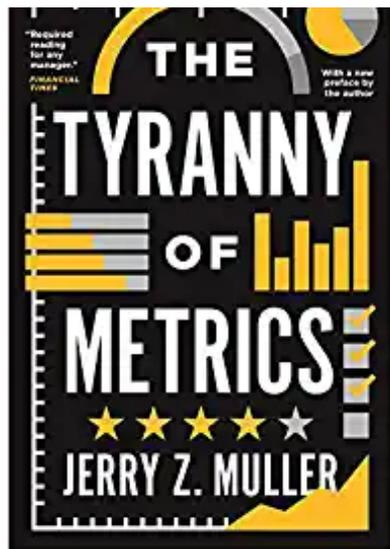
Frank Pasquale

Presentation for the ETUI

May 17, 2022

Metric Power and the Forces of Production
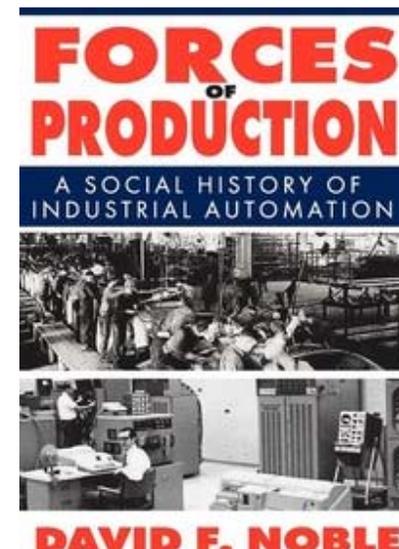
A Cross-Cutting Concern for the Future of Labor: The Digitization of Judgment

Traditional Gatekeeper	Challenged or Enhanced by Technology
Teacher decides grades in part based on which students are paying attention.	"ClassCare" system records all students' faces, analyzing expressions of attentiveness second-by-second or minute by minute, and scoring for engagement. It scores the teacher as well.
Human resources department reviews resumes and then interviews candidates.	Algorithmic resume sorter finds qualified candidates; AI reviews video interviews for likely "fit"
Suicide prevention hotline waits to receive call from a person in need or someone concerned about them.	Algorithms automatically alert Facebook employees about concerning posting behavior. See also Samaritans Project and Project Durkheim.
Police chief assigns officers to neighborhoods based on recent criminal activity.	"Predictive policing" algorithm combines past arrest data with information on weather, recent economic trends, movements of suspects, and much more.
Intensivist at ICU helps guide families toward decisions about end of life care.	"Death algorithm" compares current ICU patient with "digital doubles" and reports on median time until death and chance of recovery.

Digitization Supplementing or Replacing Professional Judgment

Traditional Gatekeeper	Challenged or Enhanced by Technological Scoring
Attorney decides whether a case is likely to succeed	NLP via ML compares draft complaint with extant library of complaints and scores for likelihood of success (similarity to past successful complaints).
Editor decides what stories should be on the front page of a newspaper	FB, Twitter, Google News, WeChat, or other algorithms score posts

Something to consider as ratings expand in health care…

| Raters Rated	1. Patients or their agents	2. Doctors	3. Hospitals	4. Insurers
A. Patients	1.A. YouTube comments on plastic surgery	2.A. Medical Justice Database	3.A. Medication adherence or compliance score	4.A Intelliscript and databases on risk
B. Doctors	1.B. Angie's List Yelp/Consumer Reviews	2.B. Physician peer review (HCQIA)	3.C. ACO dashboards	4.B. NY AG Settlement
C. Hospitals	1.C. Consumer Reports/Hospital Compare	2.C. Discussion forums	3.C. Comparative ads	4.C. Tiering and cost-sharing
D. Insurers	1.D Insurance exchanges	2.D. Payer reviews	3.D. Rankings of "worst payers"	4.D Medicare views on Medicare advantage Reinsurance

Are these appropriate tools of evaluation?

- "If lenders see political activity on someone's Twitter account in India, they'll consider repayment more difficult & not lend to that individual."

- "[One firm] analyses the way you fill in a form (in addition to what you say in the form), and how you use a website, on what kind of device, and in what location."

- Parkinson's prediction via mouse tremors.

Divorce rate in Maine per 1,000 people

Per capita consumption of margarine (lbs)

Correlation: 99%
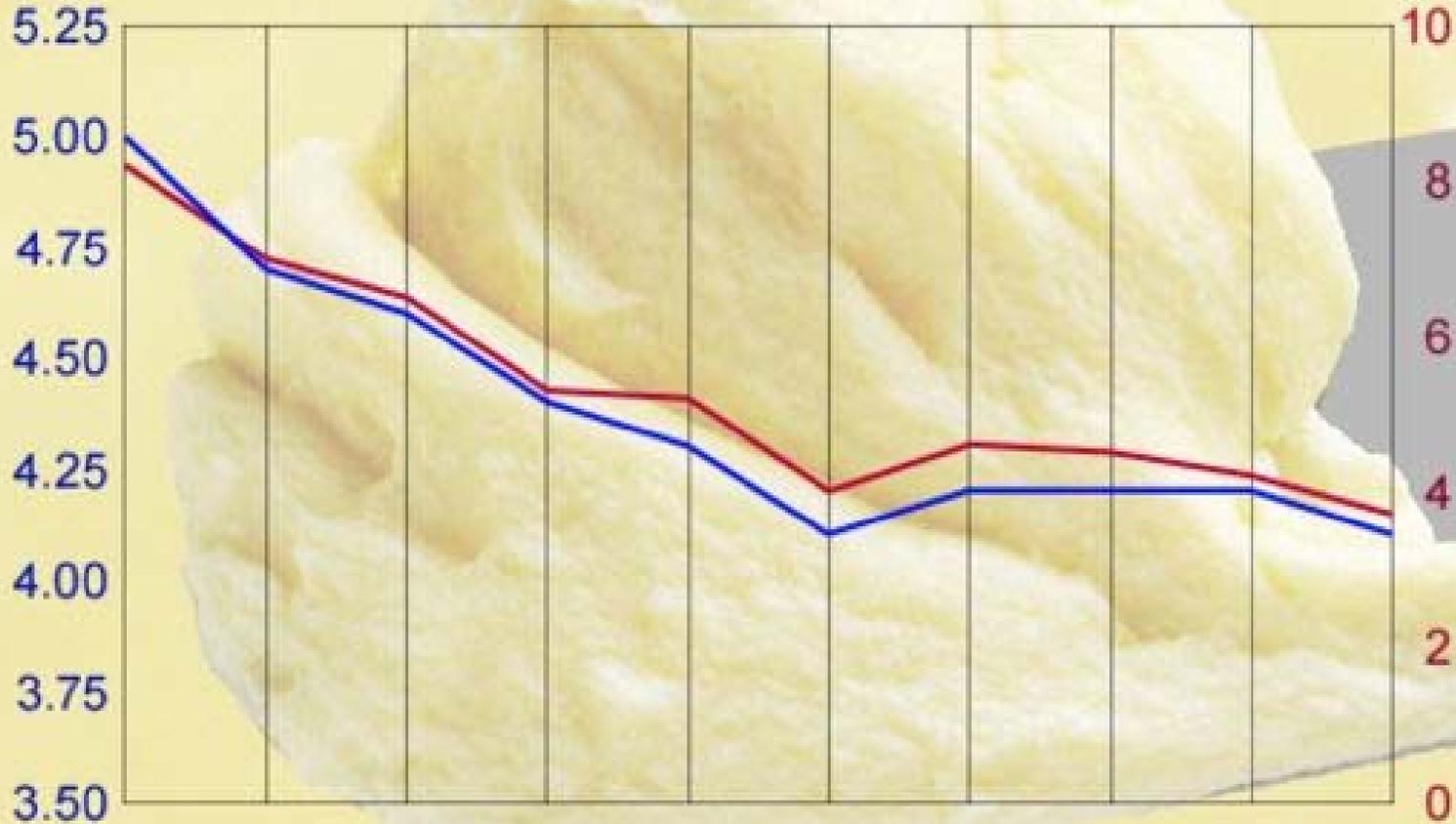
Stages of Machine Learning and Reinforcement Learning

- *Stages of Machine Learning*
  - Problem Definition
  - Data Collection
  - Data Cleaning
  - Summary Statistics Review
  - Data Partitioning
  - Model Selection
  - Model Training
    - Tuning
    - Assessment
    - Feature Selection
  - Model Deployment
    - From Lehr & Ohm, "Playing with the Data"

- Focus on Reinforcement Learning by an Agent
  - Model (describes the environment of the agent, and the agent's relationship with that environment)
  - Policy (how the agent makes decisions)
  - Reward (the goal of the agent)

Data and Bias

- Accuratism frames answer to bias as "more data"
- But critical social science questions that narrative.
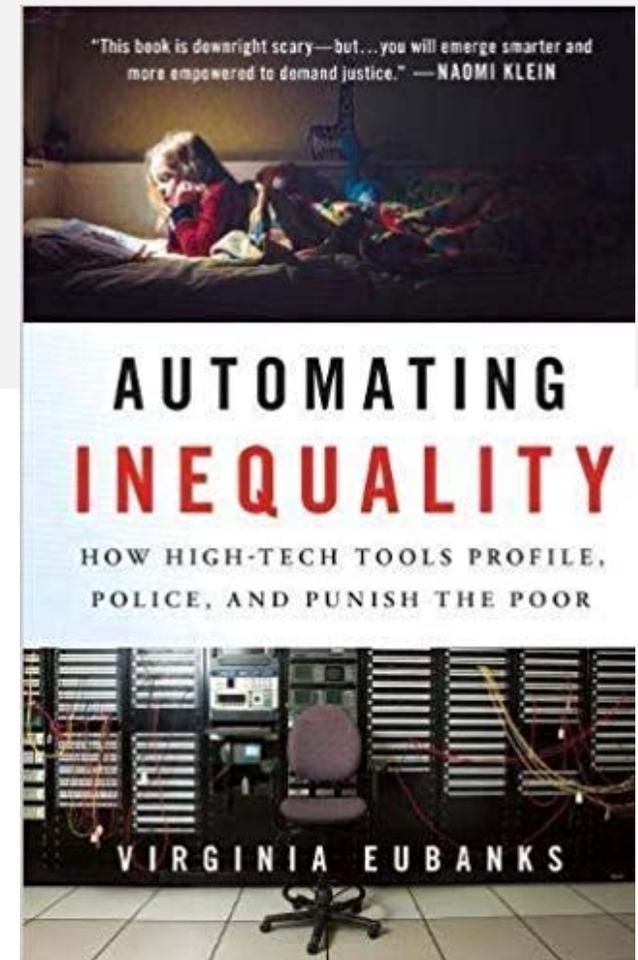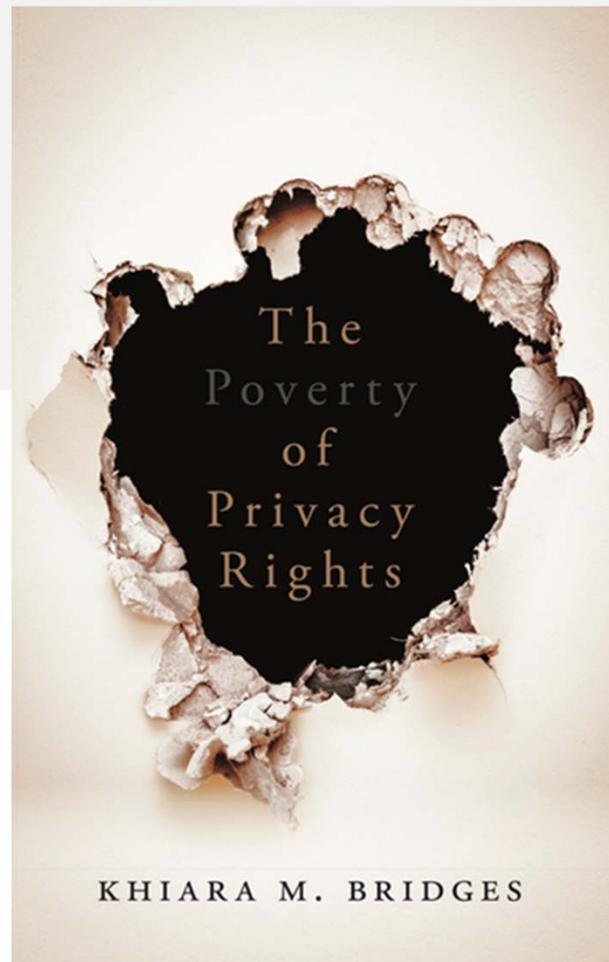
Campbell's Law

- The more any quantitative social indicator (or even some qualitative indicator) is used for social decision-making,
  - the more subject it will be to corruption pressures
  - and the more apt it will be to distort and corrupt the social processes it is intended to monitor.
- Related idea: Goodhart's Law
  - David Manheim & Scott Garrabrant, "Categorizing Variants of Goodhart's Law" (describing "several distinct failure modes for overoptimization of systems on the basis of metrics").

Preliminary Conclusion

- The case for limits
  - Surveillance that is too intensive
    - 24/7 phone location monitoring
    - All phone data
    - Note "anti-fraud exceptions" to privacy law

- The case for refinement
  - Toward more objective performance assessment
  - Possibilities for democratized metrics.
    - However:
      - Politics is inevitable
      - Campbell's/Goodhart's law will lead to corruption pressures