Watchful

The **Machine Teaching** Platform for the Enterprise

Contact John
COO & Co-Founder

Problems with **Hand-Labeling**, and the Efficacy of **Automation Techniques**

Major Issues with Hand-Labeling

Introduction of Bias

- Hand-labels are not interpretable or reproducible, and are inherently bias-prone

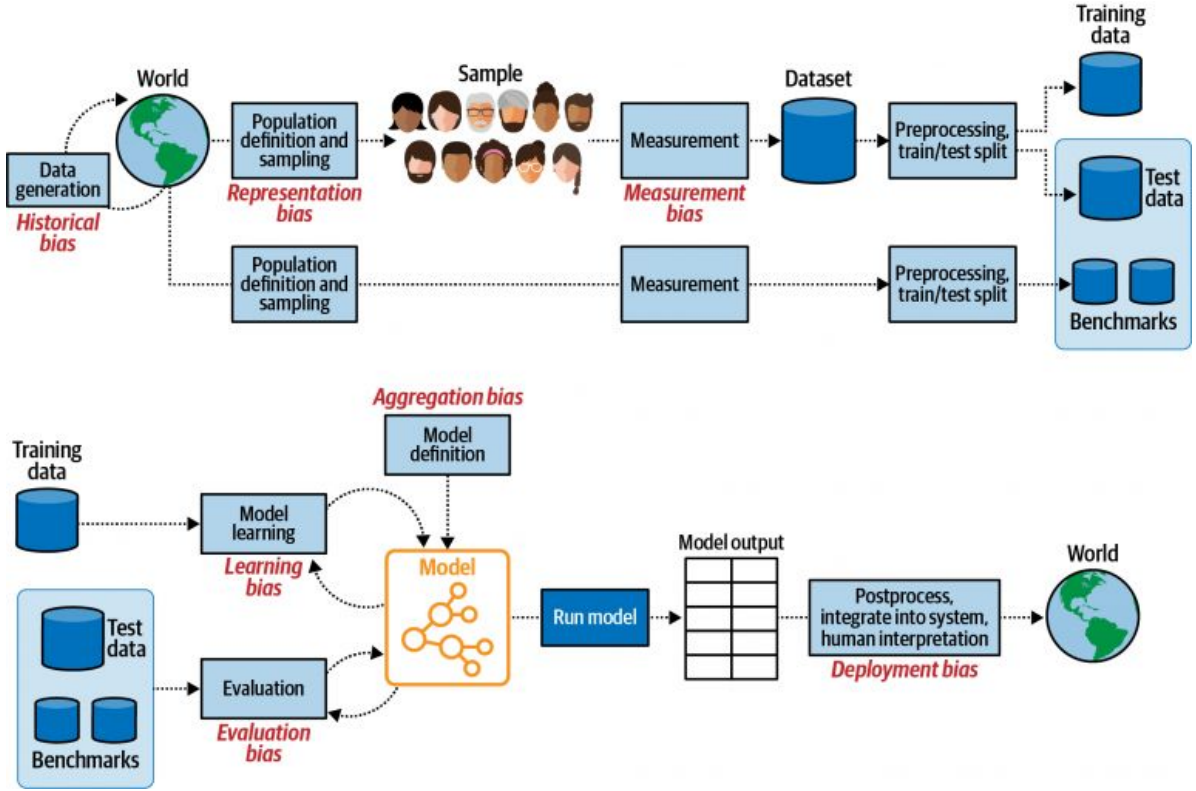
Prohibitive Costs

- Models that require lots of data or subject matter interpretation of the data are frequently cost prohibitive to build

Ground Truth Is a Lie

- The real world is full of shades of gray
- Hand-labeled data often does not capture the nuance of inter-annotator disagreement

Algorithmic Bias



Source: Hand Labeling Considered Harmful

Algorithmic Bias

- Often creeps in through data
- Can’t explain hand-labels
- Can’t easily remedy bias in hand-labels

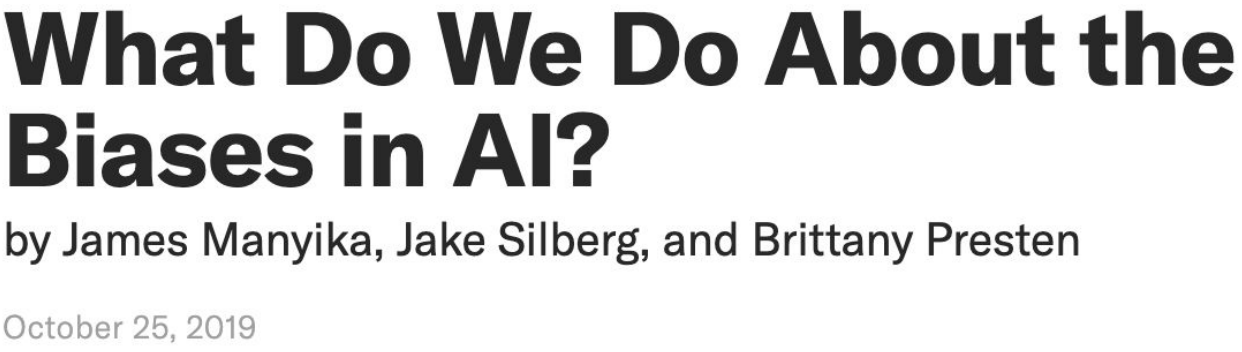
Hand-Labels: Lack of Interpretability & Explainability

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016



What Do We Do About the Biases in AI?

by James Manyika, Jake Silberg, and Brittany Presten

October 25, 2019

The World of Alternatives

Low — **Amount Supervision Required** — High

Machine Teaching
Collection of techniques to extract knowledge from humans for model training

Weak Supervision
Noisy heuristics are used to weakly label large amounts of data for machine learning

Semi-Supervised Learning
Combine a small amount of labeled data with a large amount of unlabeled data

Active Learning
Algorithm queries users interactively to label specific segments of the data

Synthetic Data Generation
Building models to generate data points that have the same statistical validity as “real” data

Transfer Learning
Leveraging general pre-trained models to quickly bootstrap specific models

Machine Teaching: Introducing Explainability

- Techniques like weak supervision offer a framework for interpretability in labels
- Often must trade interpretability for quality
- Can combine approaches to achieve the right levels of interpretability, performance, and quality

The Algorithmic Auditing Trap

'Bias audits' for discriminatory tools are a promising idea, but current approaches leave much to be desired

Mona Sloane Mar 17 · 7 min read ★

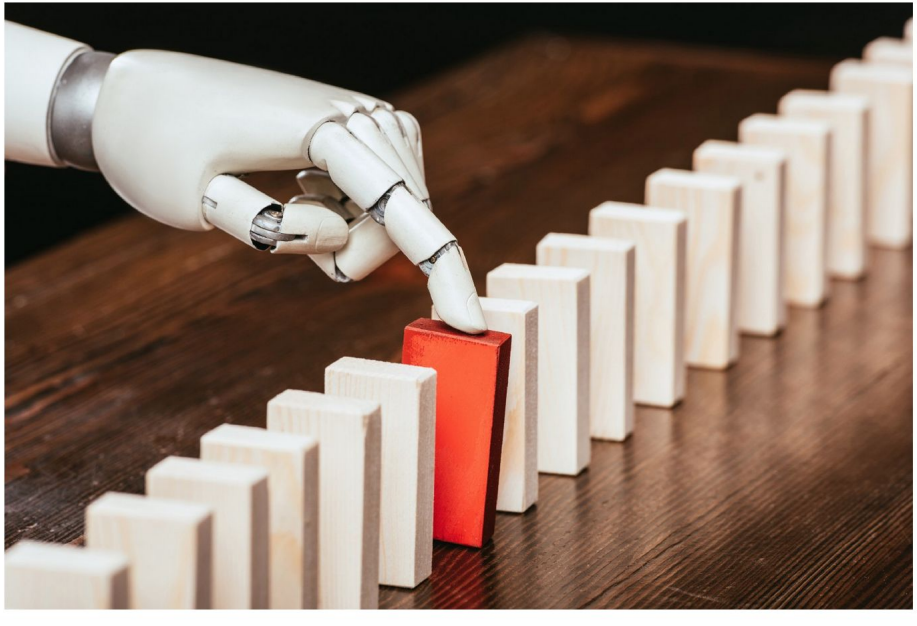


Image: LightFieldStudios/Getty Images

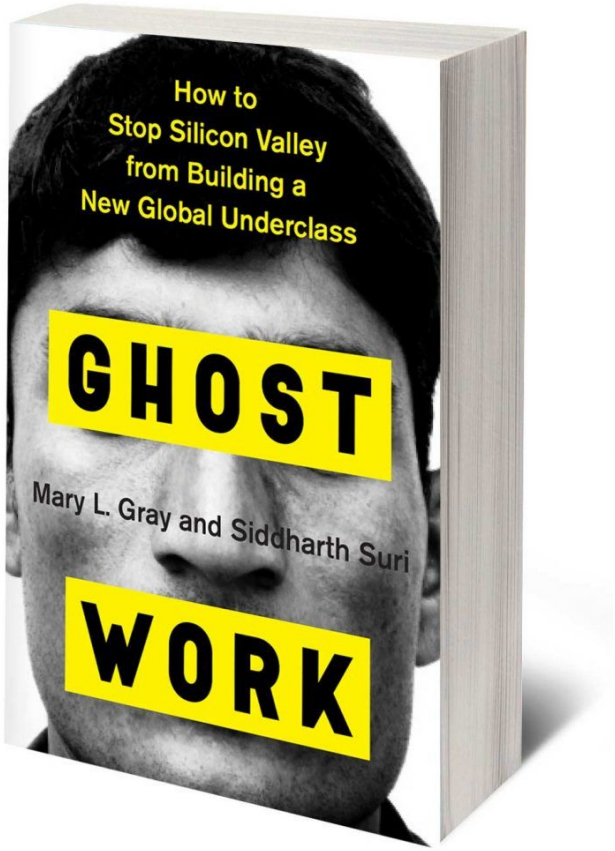
Societal, **Time**, and **Financial** Costs of **Hand-Labeling**

MOTHERBOARD
TECH BY VICE

Underpaid Workers Are Being Forced to Train Biased AI on Mechanical Turk

Workers who label images on platforms like Mechanical Turk say they're being incentivized to fall in line with their responses—or risk losing work.

AN By Aliide Naylor



Time of SMEs

- The time of experts is the scarcest resource
- You’re never done labeling
- Time spent by experts must be measured over the lifetime of the model

Time of SMEs: Not Vertical Specific

Healthcare

- Clinical Trial Matching
- Clinical Decision Support

Finance

- Fraud detection
- Contract Intelligence

Insurance

- Risk Classification
- Claims Fraud Detection

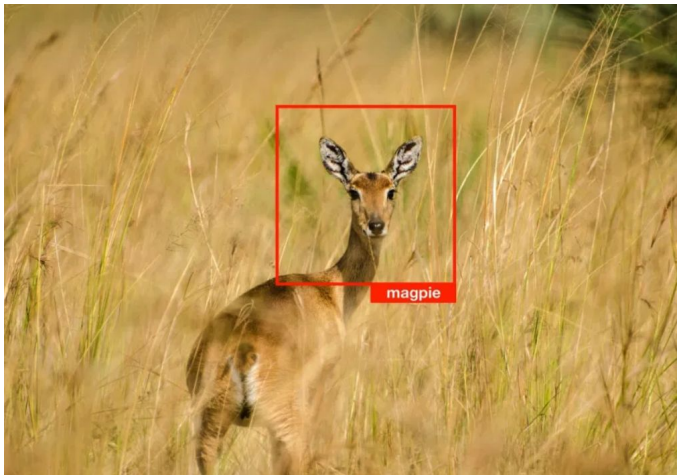
Cost to Accuracy, Reality, and Ground Truth

Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks

Curtis G. Northcutt* ChipBrain, MIT

Anish Athalye MIT

Jonas Mueller Amazon



Measuring Model Biases in the Absence of Ground Truth

Osman Aka*
Google

Ken Burke*
Google

Alex Bäuerle†
Ulm University

Christina Greer
Google

Margaret Mitchell‡

Financial Costs of Hand Labeling

- You’re never done labeling
- Class definitions often change as labeling progresses
- Cost of SME time compounds cost of overall pipeline

The Efficacy of Automation Techniques

Scaling to Very Very Large Corpora for Natural Language Disambiguation

Michele Banko and Eric Brill
Microsoft Research
1 Microsoft Way
Redmond, WA 98052 USA
{mbanko,brill}@microsoft.com

DEEP LEARNING SCALING IS PREDICTABLE, EMPIRICALLY

Joel Hestness, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md. Mostofa Ali Patwary, Yang Yang, Yanqi Zhou

{joel,sharan,ardalaninewsha,gregdiamos,junheewoo,hassankianinejad,
patwarymostofa,yangyang62,zhouyanqi}@baidu.com

Baidu Research

The Efficacy of Automation Techniques

*“We empirically validate that **DL model accuracy improves as a power-law as we grow training sets for state-of-the-art (SOTA) model architectures** in four machine learning domains: machine translation, language modeling, image processing, and speech recognition. These power-law learning curves exist across all tested domains, model architectures, optimizers, and loss functions.”*

-- Hestness et al. 2017