Matching Machine Learning Capabilities to Public Health Needs

Stephen Eubank*
BI-2026-1

March 2, 2026

* As little AI and as much teamwork with various incarnations of the Biocomplexity Institute as possible was used to generate this presentation.

My life's prompts

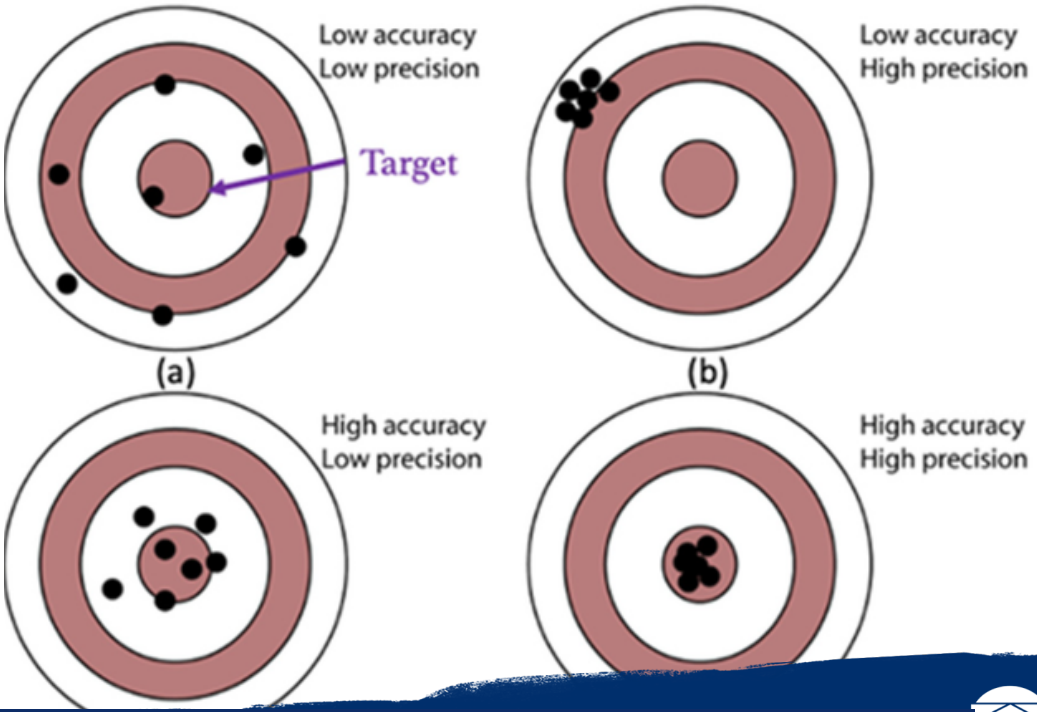
- Mathematical, statistical, physical modeling with
- Machine Learning tools:
  - time series analysis (nonlinear, nonparametric)
  - artificial neural nets (backprop & "optimal brain surgery" era)
  - decision / CART trees on what was quaintly called Big Data
  - simulation (ABM, interaction-based)
- applied to:
  - fluid dynamical and other physical systems
  - financial markets
  - natural language processing
  - transportation system planning
  - infectious disease epidemiology
- at:
  - national labs
  - startup
  - DoE, DoT, DoD, DHHS
  - State Dept. Office of Global Health Security and Diplomacy

I prefer the hypothesis testing dialect to the AI dialect

- AI:
  - hallucination
  - sycophancy
  - singularity / end of humanity
- Modeling error & statistical hypothesis testing tradeoffs
  - type [1,2] error
  - bias / variance
  - power / sample size / significance
  - precision / accuracy

Attributes of model influence test results

- Significance: complement of the probability of rejecting null hypothesis when it is true (avoiding false positives, p-value)
- Power: probability of rejecting the null hypothesis when the alternative hypothesis is true (avoiding false negatives)
- Bias: roughly, complement of accuracy
- Variance: sensitivity to variations in training set (roughly, precision)



Eagleson et al., Verification, Evaluation, and Validation, Journal of Imaging, **9**, 20 (2023).

Rules of thumb are not magically outmoded by AI

- Power / sample size / significance tradeoffs
  - correcting for multiple testing $\Longrightarrow$ decreasing power
  - decreasing effect size $\Longrightarrow$ decreasing power
  - increasing variability in null $\Longrightarrow$ decreasing power
- Bias / variance tradeoff
  - increasing model flexibility $\Longrightarrow$ less bias but more variance
- Curse of dimensionality

  sample size must increase exponentially with dimension

  (dimension $\sim$ # parameters $\sim$ model flexibility)
- Interpolation is safer than extrapolation (generalization)
- Garbage in, garbage out
  - especially relevant for LLMs trained on "the internet"!!

Optimization by exhaustive search

Training a model with

- a finite set of discrete-valued parameters
- on exa- / zetta- / yotta-bytes of data
- investing exponential amounts of computing power

is, in effect, making an exponentially large number of comparisons.

AI models are almost certainly overfit and/or data starved.

A cautionary tale of power

30 years ago, it was possible to build,

- on a Sun SPARC workstation,
- in a day or two,
- a "two-neuron" (i.e., Occam-approved) gold price model
- that performed well on 10 years of daily, de-trended data
- and performed well on 10 years of out-of-sample data
- but failed horribly on 10 years of test data

I'm not convinced that corrections for multiple testing are sufficient.

AI Weaknesses

Avoid relying on AI in domains with

- limited data
- poorly understood confounders
- complex multivariate interactions (i.e., weak univariate effects)
- undeveloped alternative hypotheses

Sounds like almost any public health problem!

AI Strengths

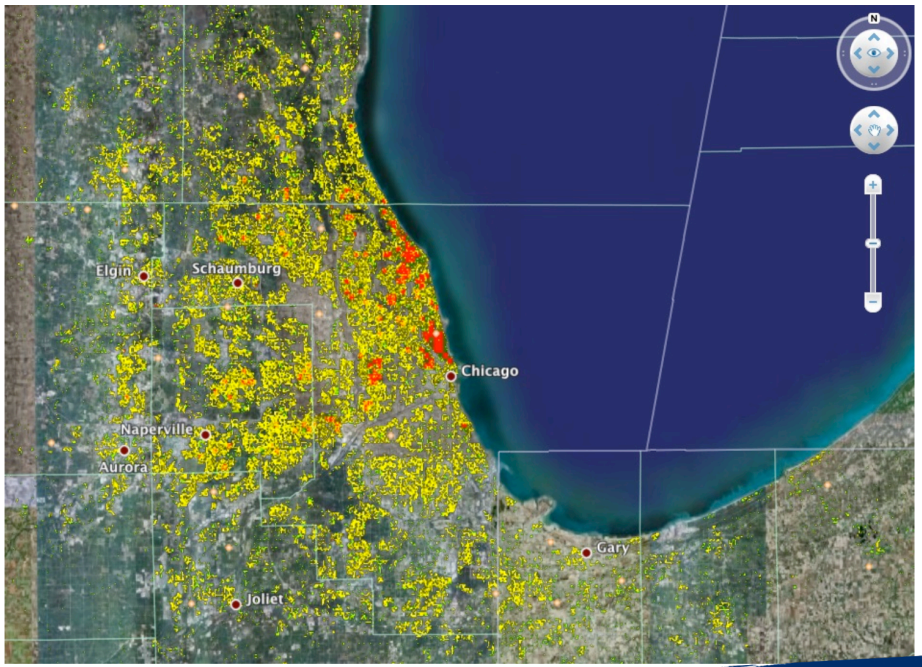
Supervised AI is well-suited for:

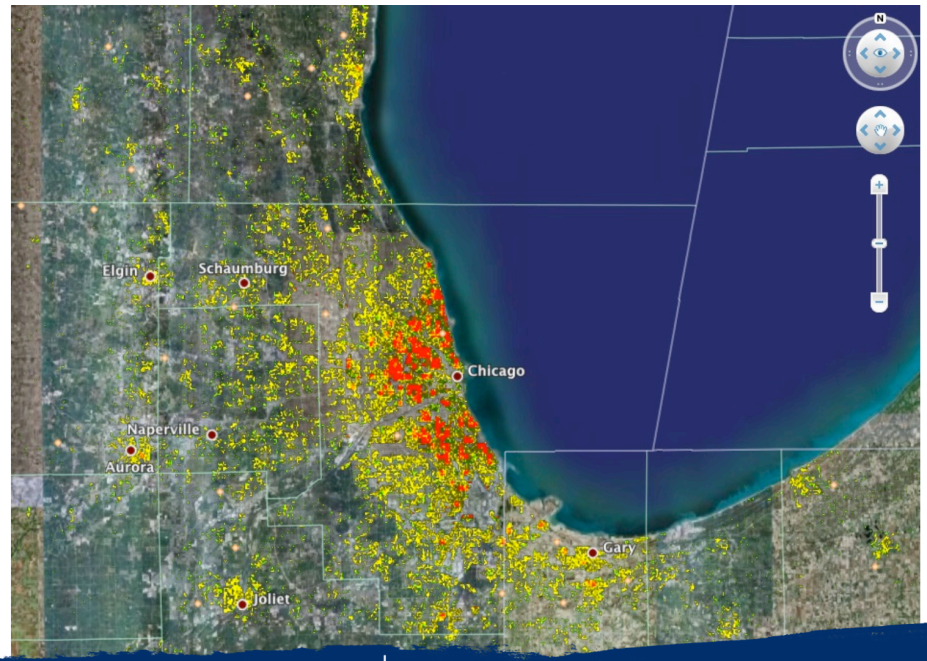
- well-specified, rote, tedious or complicated tasks, e.g.
  - disambiguating line lists
  - data fusion
- in-depth research into published work;
- tasks whose solutions can be tested;
- deducing the implications of assumptions.

Treat AI like you would a master's student.
A master's is about learning what's already known;
a PhD, about discovering new things.

Previous ML-assisted health policy informatics

- locating field hospitals for best accessibility
- predicting whether specific hospitals will reach capacity
- evaluating cohorting at training bases
- staging scarce treatment resources
- exploring HIV infrastructure needs with NIH
- evaluating equity of hypothetical policies





General categories that seem ripe for AI solutions

- data cleaning, harmonization, fusion, and analysis
- conducting systematic reviews (The Journal of AI Meta-analyses???)
- situation assessment (nowcasting instead of forecasting)
- pattern recognition
- design and statistical analysis of experiments
- public-facing natural language interfaces

Many PH policy needs require extrapolation/generalization

- identifying transmission hot spots
- estimating transmission networks
- drug discovery
- identifying subtle inequities of policies
- predicting incidence in small areas
- understanding moral hazard of PREP

Maybe informing policy isn't the right target

"If the AI debate stays focused on optimizing white-collar productivity, it will miss the more urgent opportunity: using AI to support, sustain and elevate the people that our economies need and increasingly can't find."
– Sander van 't Noordende, CEO, Randstad*

* Ranstad is one of the world's two largest multinational human resource consulting firms.

Supporting public health professionals

I'm here to listen, but will suggest a few possibilities:

- specialized chat-bots
  - triage on phones or at clinics
  - translating / writing locally appropriate messaging
  - providing pseudo-patients for training therapists
  - discussing stigmatized behaviors (surveys, treatment support, etc.)
- reducing [non-HIV] burdens on time
- suggesting novel uses for approved drugs

All of these would require carefully-designed safety and efficacy trials before supervised use.

Thank you! Here's a present for listening to the end

The "spurious correlations" website is back, now with AI-generated explanations and pre-written academic papers, including the prompts that generated them:

https://tylervigen.com