

# Troubling Trends in Machine Learning Scholarship

Zachary Chase Lipton  
Carnegie Mellon University  
Amazon AI

Email: [zlipton@cmu.edu](mailto:zlipton@cmu.edu)

Shenanigans: [@zacharylipton](https://twitter.com/zacharylipton)

<https://arxiv.org/abs/1807.03341>

# Alchemy or Science?

## The metaphor

- What sort of science is machine learning in the first place?
- Are we confident that alchemy was not science?

## Slippage of rigor

- What is really happening?
- Why is it happening?
- What are mechanisms costs of pushing back?

# Pre-paradigmatic Science

- Immature stage of inquiry
- Try random things



# Normal Science

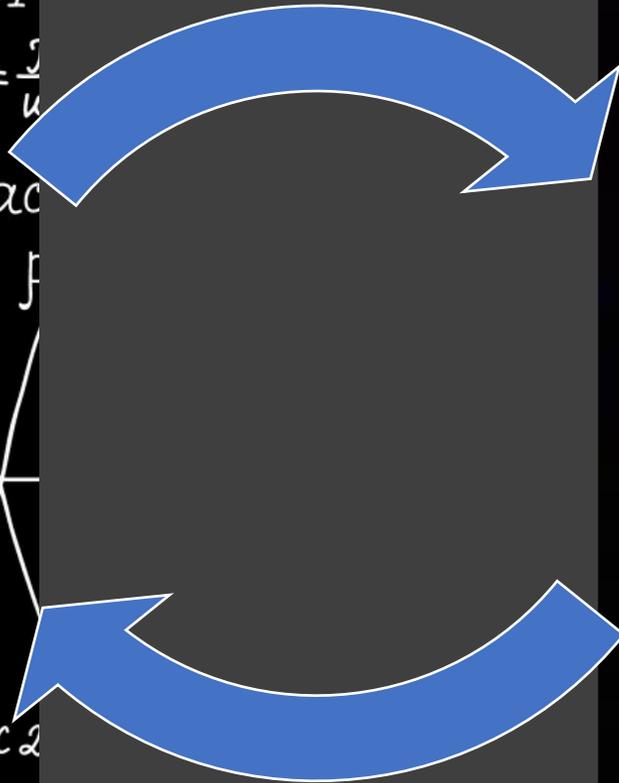
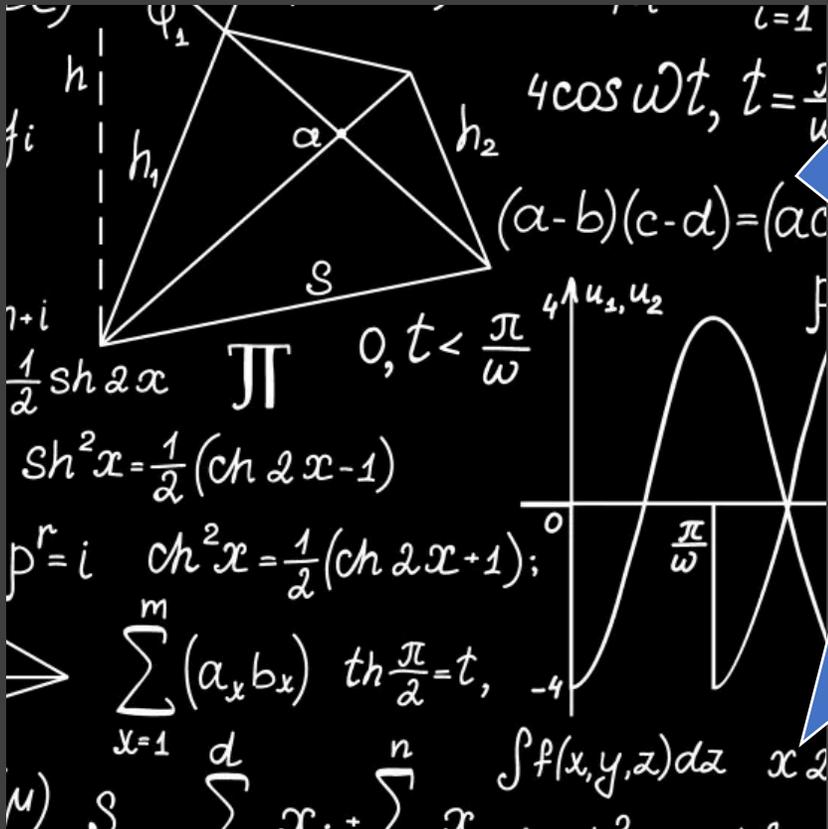
- Takes place upon *acquisition of a paradigm*
- Works within a rigid set of assumptions
- Exhibits a *peculiar efficiency*
- Tends towards *puzzle solving*



# Paradigm Shift

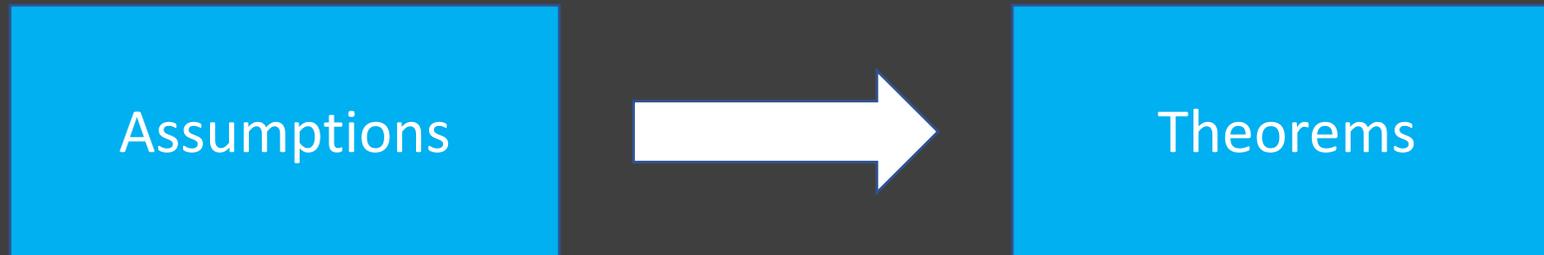
- Normal science encounters problems it cannot solve / anomalies it cannot explain
- Paradigm shifts are the “tradition-shattering complement to the tradition-bound activities of normal science”

# Aligning competing theories



# Machine Learning Theory

- Assumptions made but not interrogated
- Knowledge created via derivation from first principles
- Exhibits aspects of normal science



# Machine Learning Practice

- We run lots of “experiments”
- But few aimed at generating knowledge of the natural world
- Most empirical work dedicated to leaderboard chasing
- Exhibits aspects of the pre-paradigmatic

#	Δrank	Team Name <small>* in the money</small>	Score <small>?</small>	Entries
1	—	Marios & Gert <small>👤 *</small>	0.62703	198
2	↑1	The Slippery Appraisals <small>👤 *</small>	0.62450	472
3	↓1	zzspar <small>*</small>	0.62422	400
4	↑2	B Yang <small>*</small>	0.62259	93
5		baile	0.62134	173



What's so bad about  
the alchemists?

# What's so bad about the alchemists?

## The objective

- They sought “the elixir of life”
- AI/ML researchers seek “intelligence”
- Philosophy studies “existence, knowledge, values, reason, mind”

## The practice

- Played a significant role in early science
- Some of their ideas became obsolete
- Same true for Ptolemy, was he not a scientist?

# What did the alchemists get wrong?

- “...they guarded their work in secrecy including cyphers and cryptic symbolism”





# Cautionary Tale

# Journalism pre-1990

- Subscription driven readership
- Long-term careers & rewards
- Reputations forged over decades
- Consequences for botching stories
- Articles written over days or weeks



# (too much) journalism today

- Key metrics: clicks & impressions
- A la carte consumption model
- Demand-driven news production

**FACEBOOK'S ARTIFICIAL INTELLIGENCE  
ROBOTS SHUT DOWN AFTER THEY  
START TALKING TO EACH OTHER IN  
THEIR OWN LANGUAGE**

'you i i i everything else'

Andrew Griffin |  
Monday 31 July 2017 17:10 | 91 comments | 1990 shares

     Like Click to follow  
The Independent Tech

**TECHNOLOGY**

## An Artificial Intelligence Developed Its Own Non-Human Language

When Facebook designed chatbots to negotiate with one another, the bots made up their own way of communicating.

ADRIENNE LAFRANCE JUN 15, 2017



# Closer to home



# How to fix it? Who to fix it?

- Recognizing the problem far easier than fixing it
- All the knee-jerk “fixes” sound horrible
  - Censor bloggers?
  - Censor social media?
  - Nationalize the press?



Troubling Trends in ML  
(<https://arxiv.org/abs/1807.03341>)

# Artificial intelligence today



## People



Siraj Raval

Tweets  
2,790

Following  
292

Followers  
48.2K

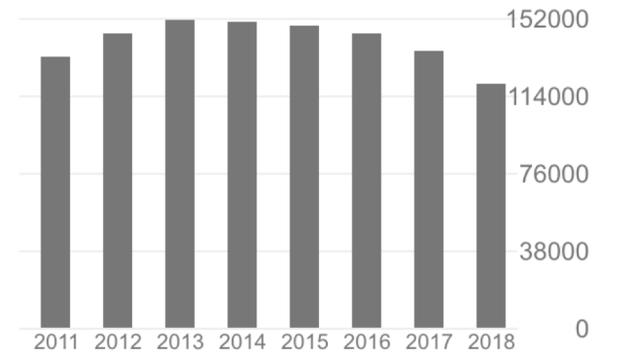
**Siraj Raval**

@sirajraval

I live to serve AI Wizards. Director of School of AI. Youtube Personality.  
Author. Subscribe! [youtube.com/c/SirajRaval](https://youtube.com/c/SirajRaval)

San Francisco, CA • [youtube.com/c/SirajRaval](https://youtube.com/c/SirajRaval)

	All	Since 2013
Citations	2856062	853173
h-index	333	275
i10-index	333	333



# Why are we here?

- Create foundational knowledge
- Prod open research questions
- Bring theory and nature into alignment
- Pose new problems to challenge scientific paradigms
- To educate readers

# The good & the bad

- +++
  - A **vibrant economy** around technologies reduced to practice
  - Financial resources for **salaries, data, and compute**
  - More of us can now pursue **full-time research careers** past PhDs
  - Strong tooling and abundant tutorials **widen accessibility** to ML
- ---
  - Overburdened **reviewers**
  - Emergence of new **incentives unaligned** with scholarship
  - Whole areas of research and discourse that are **flatly wrong**
  - Emergence of **bad academic patterns** sans consequences

# Explanation vs Speculation

- Many papers declare speculative stories as facts
- Confuse readers about open vs closed questions
- Intuition has role, clearly differentiated from evidence-based conclusions



# Un-scientific experiments

- Neglect of falsifiability
- Failure to identify source of empirical gains
- Propose  $n$  tweaks, where only  $m \ll n$  needed



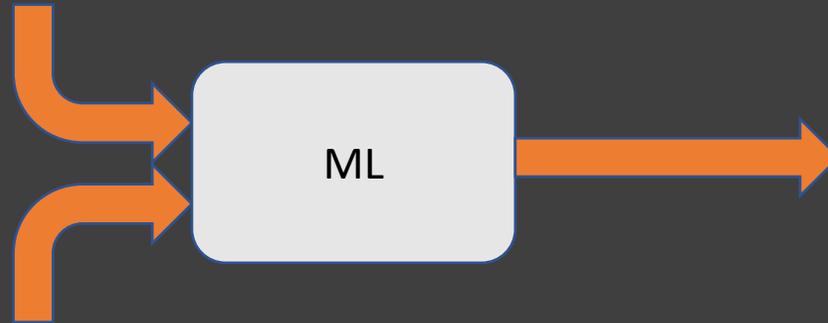
# Recent comeuppances in bad empirical ML

- Reproducibility crisis in RL ([Henderson et al 2017](#))
- Reported gains over LSTM on LM due to hyperparameter tuning ([Melis, Blunsom, Dyer 2017](#))
- Hypothesis only baselines surprisingly strong in NLI ([Poliak et al 2018](#))



# How much reading does reading comprehension require?

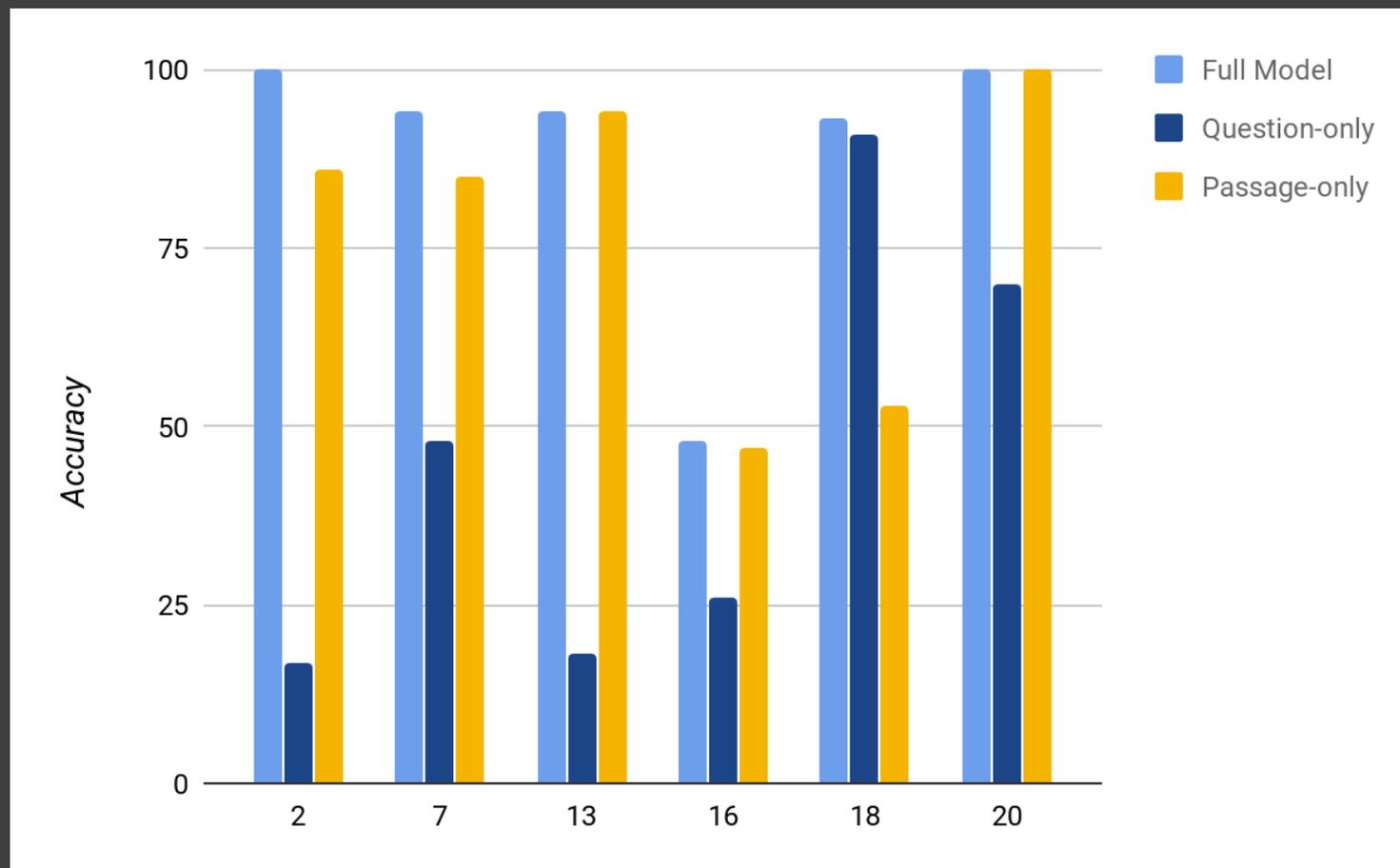
Which team has won the most Super Bowl titles?



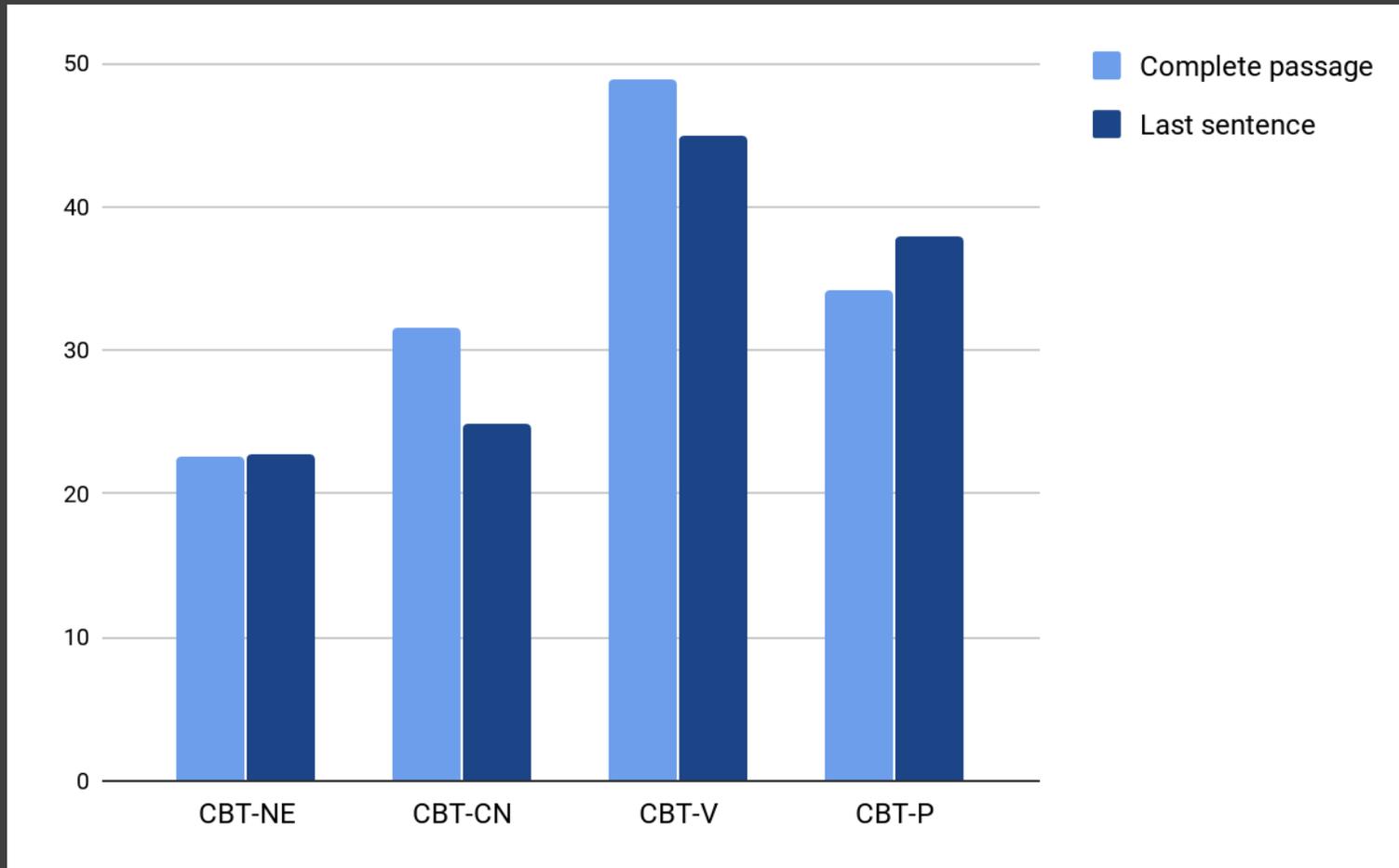
The **Pittsburgh Steelers** have the most Super Bowl championship titles, with six. The **New England Patriots** have the most Super Bowl appearances, with ten. **Charles Haley** and **Tom Brady** both have five Super Bowl rings, which is the record for the most rings won by a single player

The **Pittsburgh Steelers** have the most Super Bowl championship titles, with six. The **New England Patriots** have the most Super Bowl appearances, with ten. **Charles Haley** and **Tom Brady** both have five Super Bowl rings, which is the record for the most rings won by a single player

# Results on “AI-Complete QA” (BaBI)

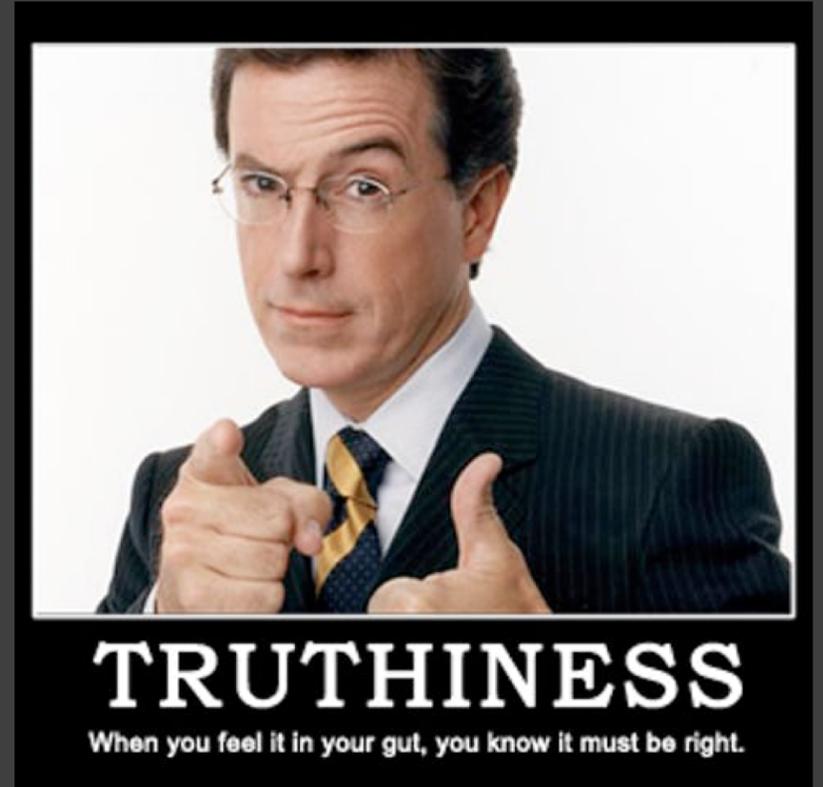


# Results on CBT

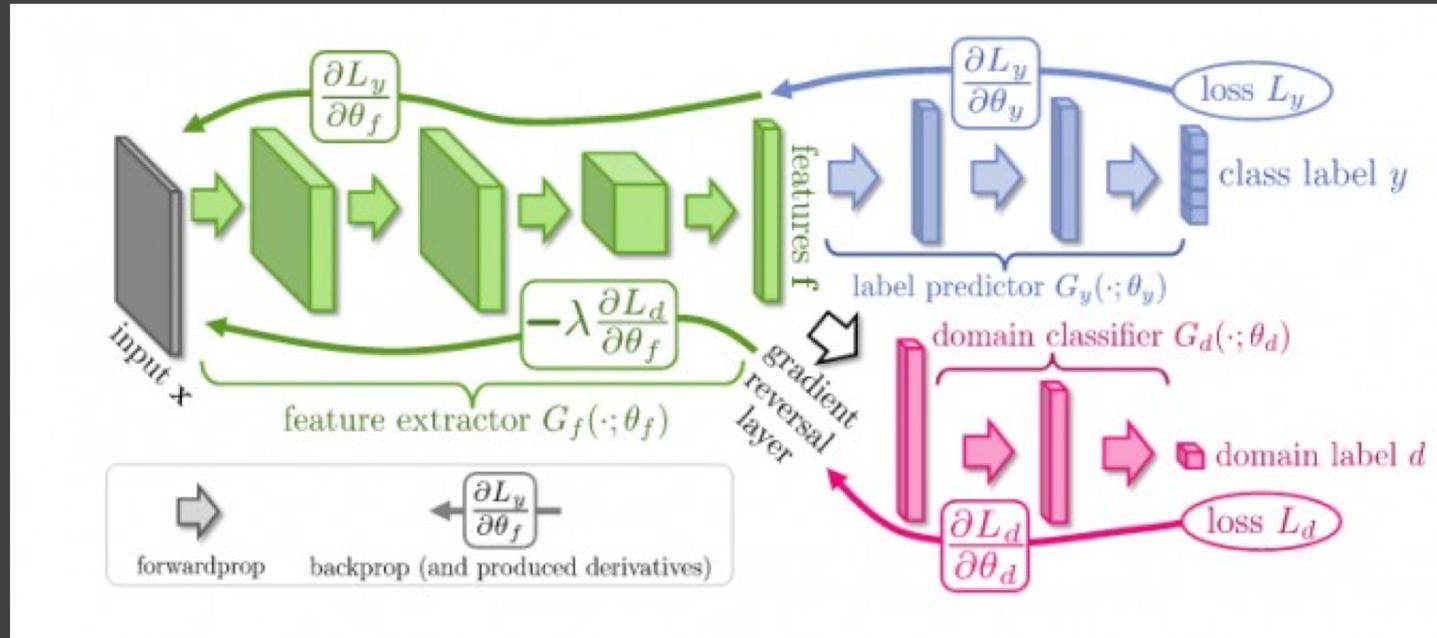


# Mathiness

- Spurious math to display *technical depth* rather than **clarity**
- **Spurious theorems** with no real connection
- **Slippage** between informal and formal terms
- Excessive notation for **steamrolling** reviewers



# Deep Domain Adaptation



$$R_{\mathcal{D}_T}(\eta) \leq R_S(\eta) + \sqrt{\frac{4}{n} \left( d \log \frac{2\epsilon n}{d} + \log \frac{4}{\delta} \right)} + \hat{d}_{\mathcal{H}}(S, T) + 4\sqrt{\frac{1}{n} \left( d \log \frac{2n}{d} + \log \frac{4}{\delta} \right)} + \beta$$

# Abuse of language

- Suggestive language
  - thought vectors, consciousness priors, reading comprehension, curiosity, fear, semantic \_\_\_\_\_
- Overloaded Technical Language
  - deconvolution, deduction, generative models, overfitting, generalization
- Suitcase words
  - fairness, interpretability, consciousness



critiquing

correcting

# Who should do what?

- Who can/should enact change?
- What variables can we intervene on?
- Individual change is hard but doable, community level change a bigger puzzle



# We want it all!

- We want financial resources, & objectives aligned w intellectual work
- We want to grow labs past bandwidth and to provide mentorship!
- We want fame and notoriety but not for the press to distort incentives
- We want strong peer review but not to become conservative like economics or statistics
- We want better peer reviewers, but culling either the community or the reviewer pool seems problematic



# What we can do

- Engage in critical, scientific, and foundational research
- Set incentives for doing ambitious work
- Protect the “R” word
  - Push back on the communities suggesting publishing is either/or  
*(while doing something short of research)*
  - Remove the incentives for spamming conferences
  - Help the press to do their job (speak candidly & honestly, on the record)
- Create tiers for publication
  - Perhaps retrospective journals can cherry pick useful papers for shepherding into more authoritative form
- Create better incentives to synthesize and distill a noisy literature