The Practice of Machine Learning

Peter Norvig

Google
Machine Learning

Standard ML Model

1. Labelled Data cat: 

Standard ML Model

1. Labelled Data — cat: 
2. Deep Learning — TensorFlow

Standard ML Model

1. Labelled Data — cat: 
2. Deep Learning — TensorFlow
3. Profit! — $$$

Standard ML Model

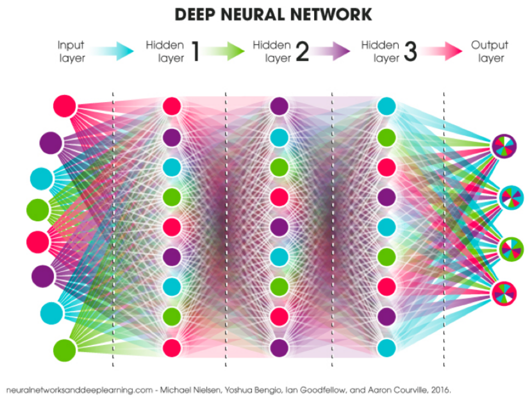
1. Labelled Data — cat: 
2. Deep Learning — TensorFlow
3. Profit! — $$$ (or Science!)

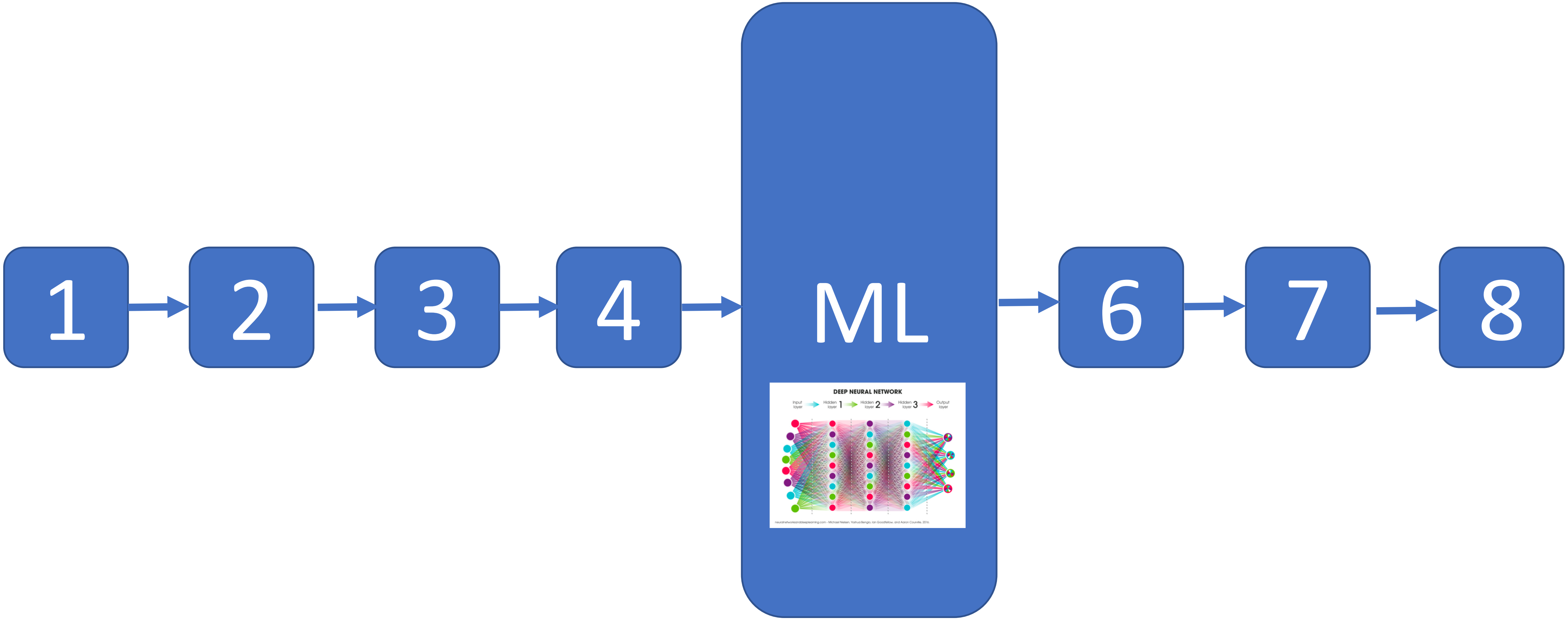
Realistic ML Model

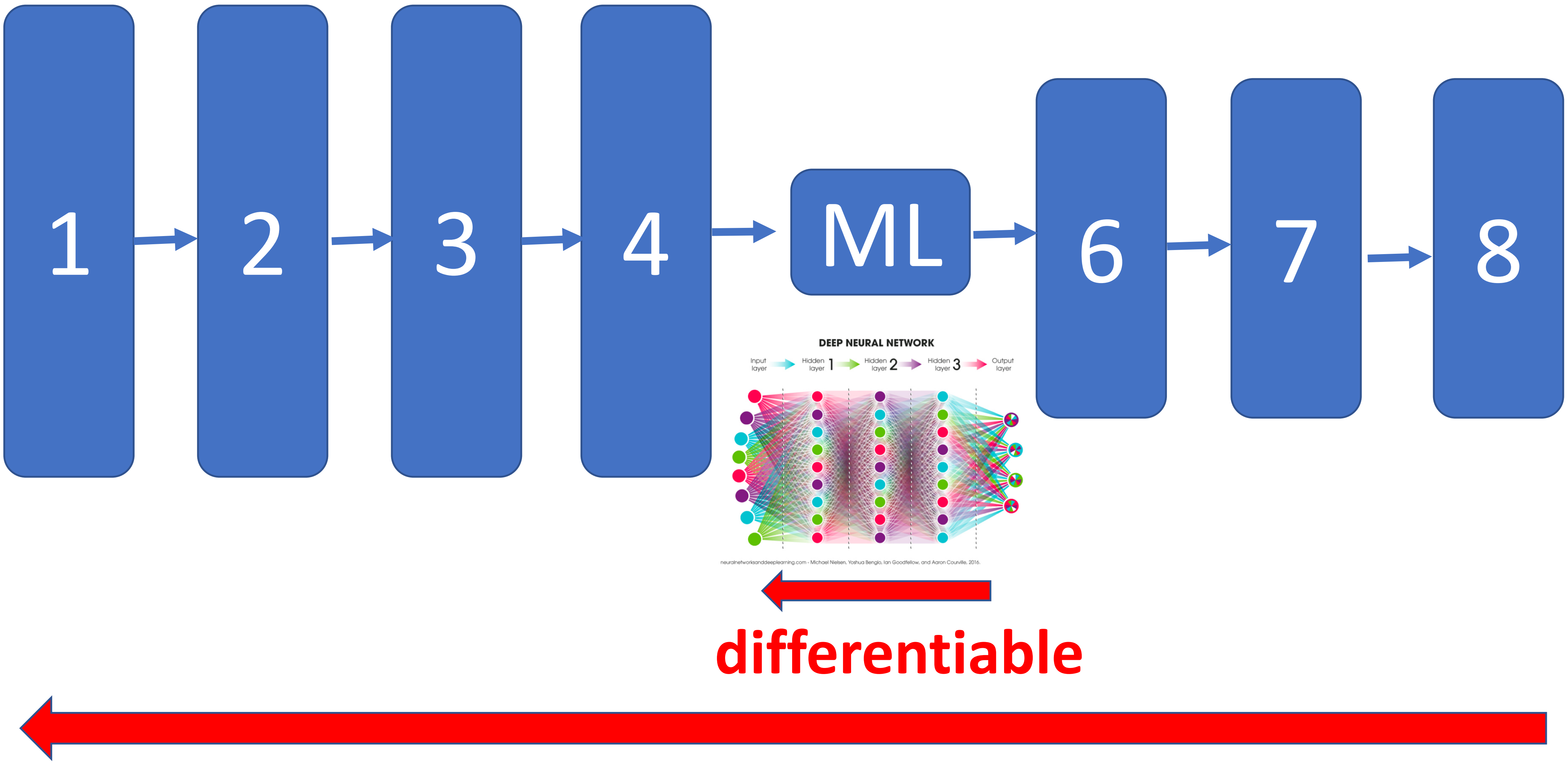
1) Explore kinds of data
2) Identify data sources
3) Curate data
4) Supervise data
5) Evaluate and debug / modify models
6) Adapt to business needs
7) Deploy, serve, monitor
8) Continually repeat: modify and improve

Realistic ML Model

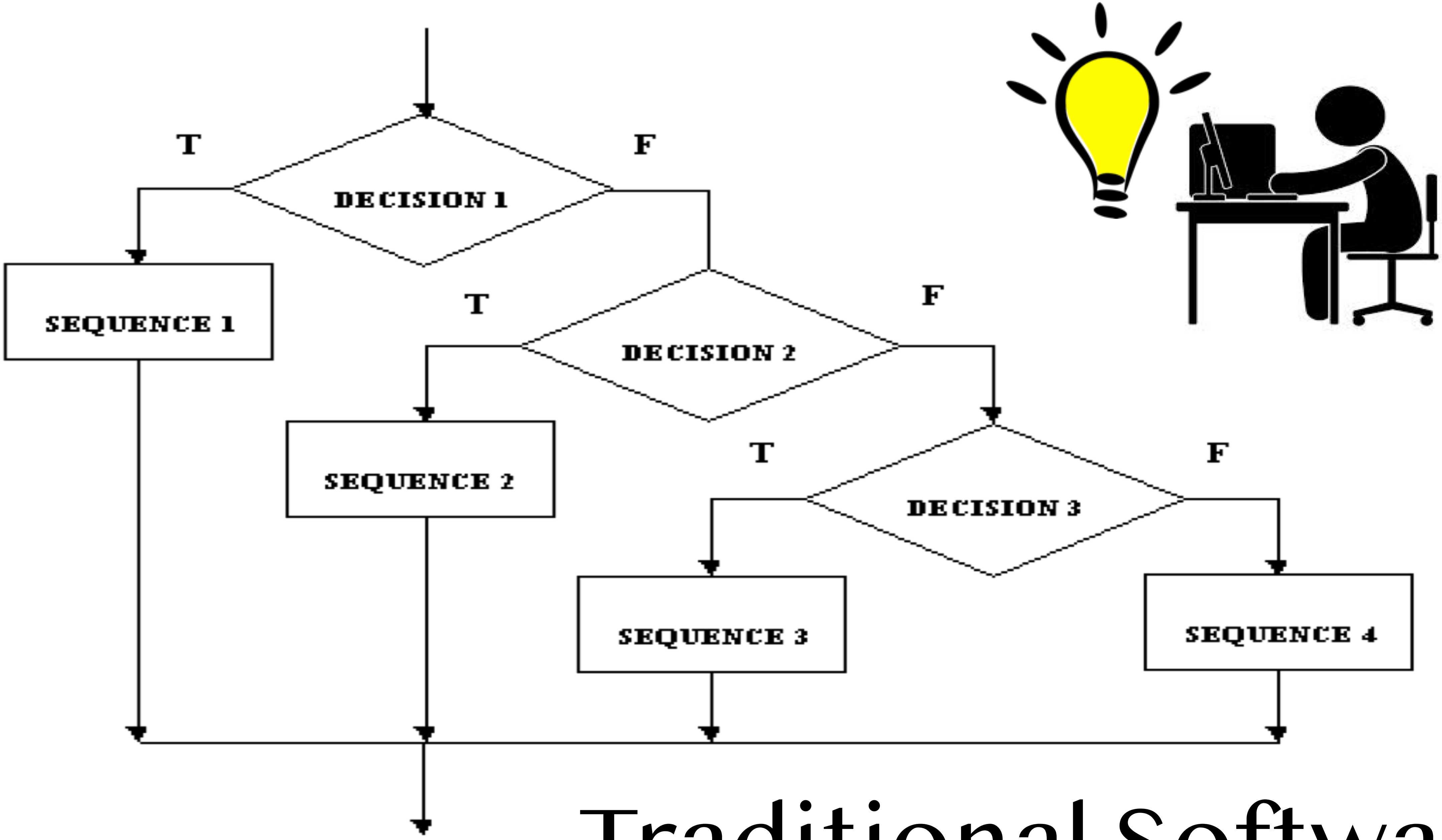
1) Explore kinds of data
2) Identify data sources
3) Curate data
4) Supervise data
5) **Evaluate and debug / modify models**
6) Adapt to business needs
7) Deploy, serve, monitor
8) Continually repeat: modify and improve



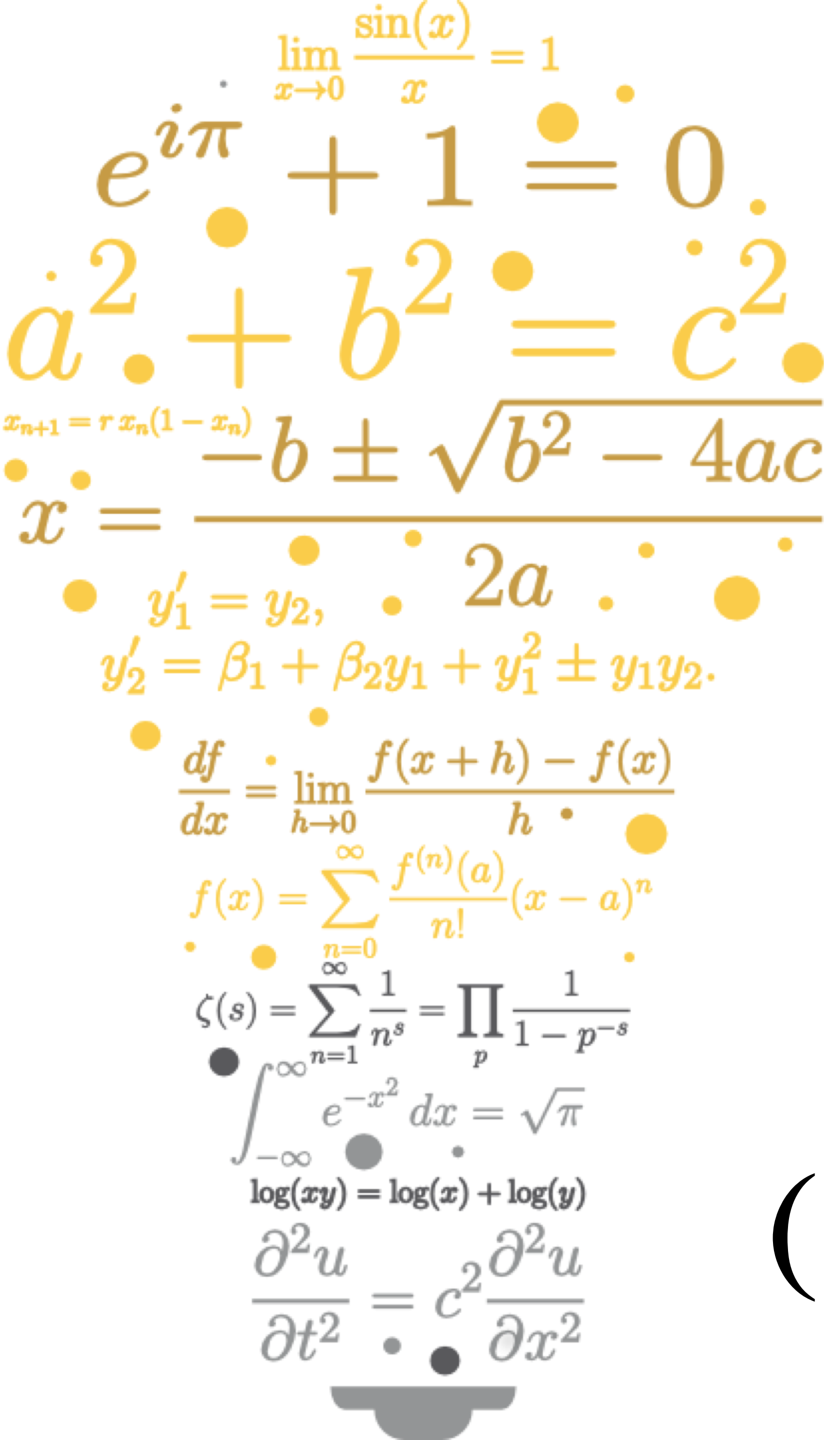

1
2
3
4
ML
DEEP NEURAL NETWORK
Input layer
Hidden layer 1
Hidden layer 2
Hidden layer 3
Output layer
neuralnetworksanddeeplearning.com - Michael Nielsen, Yoshua Bengio, Ian Goodfellow, and Aaron Courville, 2016.
6
7
8


1
2
3
4
ML
6
7
8
DEEP NEURAL NETWORK
Input layer
Hidden layer 1
Hidden layer 2
Hidden layer 3
Output layer
neuralnetworksanddeeplearning.com - Michael Nielsen, Yoshua Bengio, Ian Goodfellow, and Aaron Courville, 2016.
differentiable

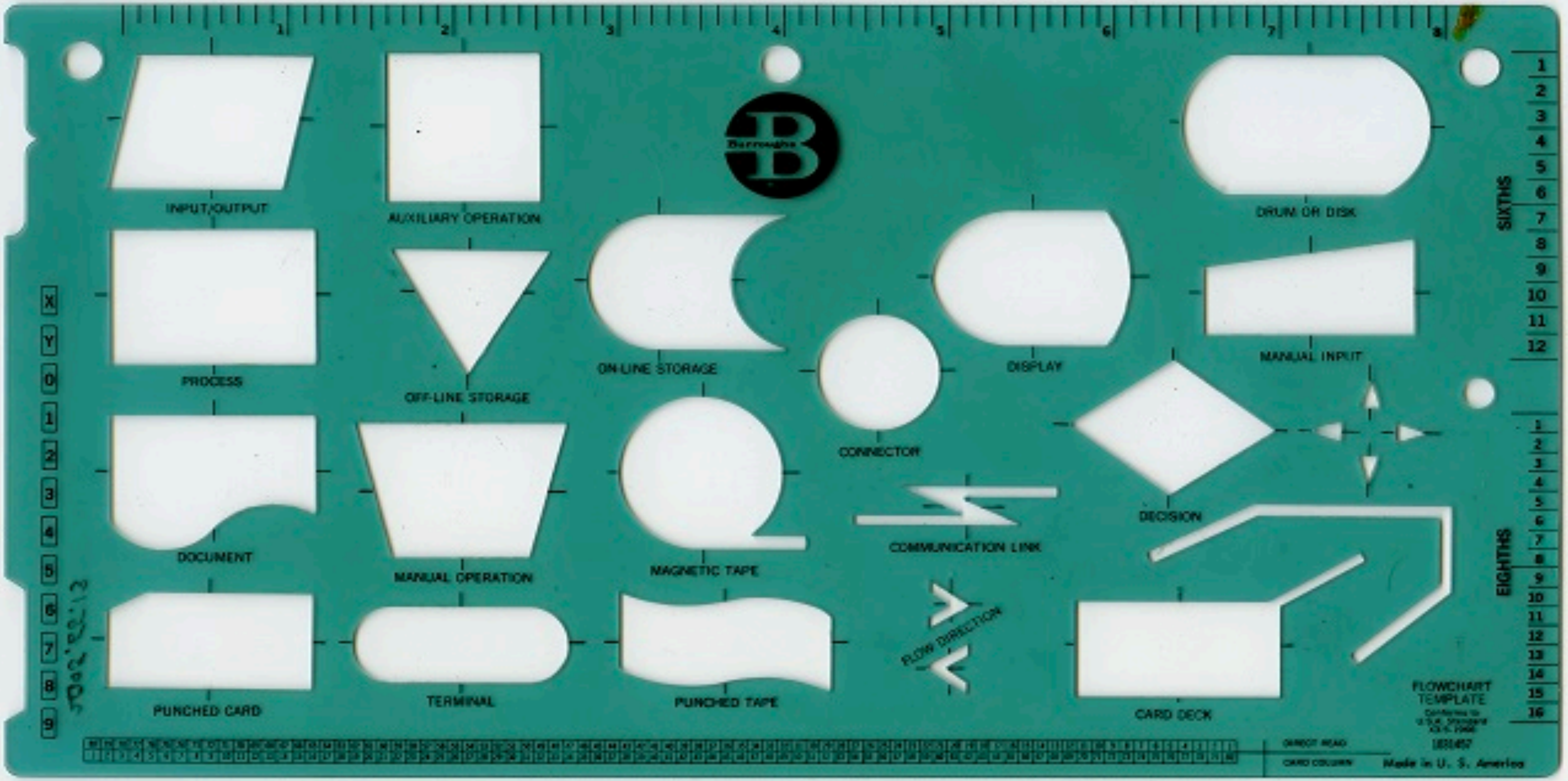
THE SOFTWARE INDUSTRY



Traditional Software



Software = Mathematical Science (Logical, Certain)



Burroughs
INPUT/OUTPUT
AUXILIARY OPERATION
DRUM OR DISK
PROCESS
OFF-LINE STORAGE
ON-LINE STORAGE
DISPLAY
MANUAL INPUT
CONNECTOR
DOCUMENT
MANUAL OPERATION
MAGNETIC TAPE
COMMUNICATION LINK
DECISION
PUNCHED CARD
TERMINAL
PUNCHED TAPE
FLOW DIRECTION
CARD DECK
SIXTHS
EIGHTHS
FLOWCHART
TEMPLATE
DIRECT READ
CARD COLUMN
Made in U. S. America

SOFTWARE DEVELOPMENT TOOLS

TRADITIONAL
PROGRAMMER = MICROMANAGER

THE MACHINE LEARNING INDUSTRY


Yoshua Bengio

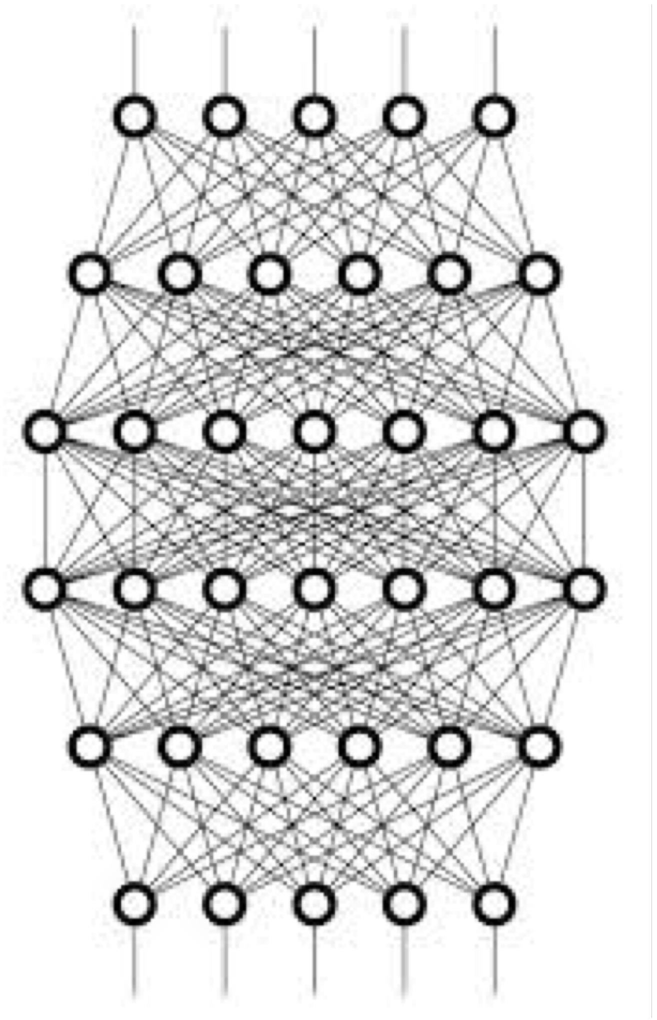
Geoffrey Hinton

Yann LeCun

Fei-Fei Li

Daphne Koller

Quoc Le



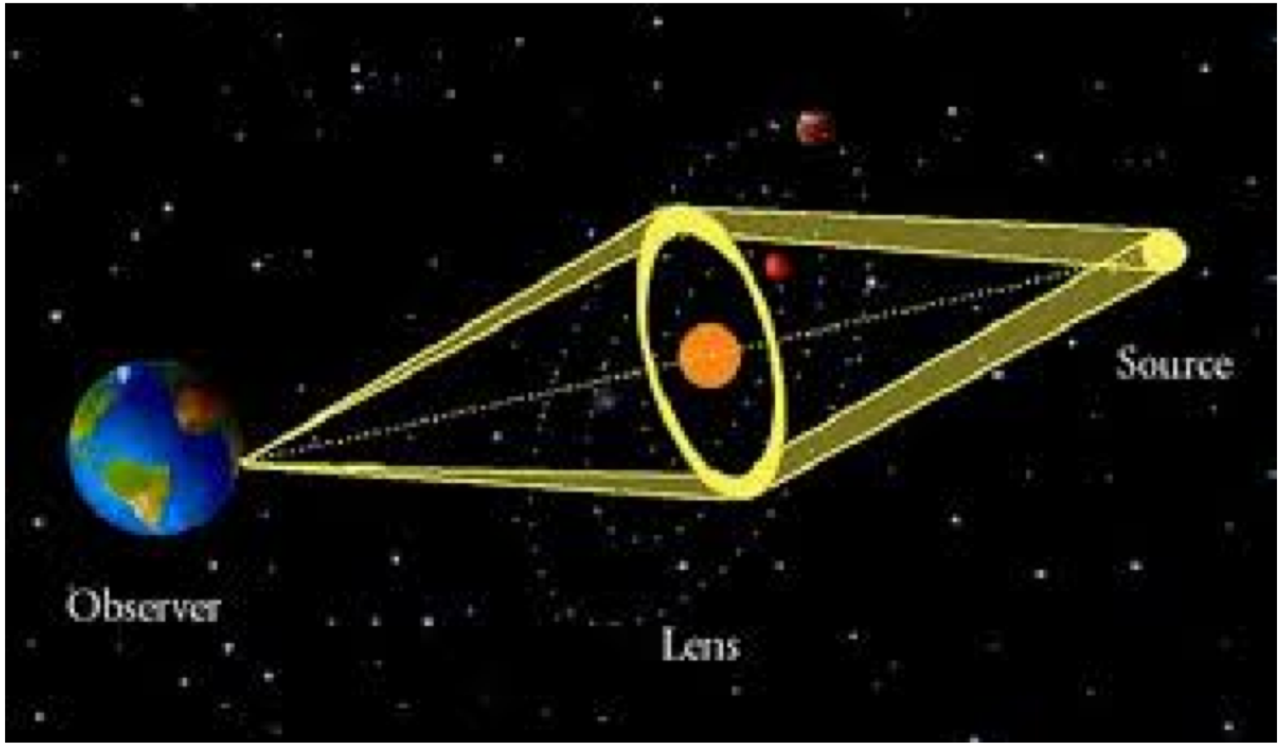
Machine Learning

Software =
Empirical
Science
(Probabilistic, Uncertain)

TensorFlow
AWS
scikit learn
K
PYTORCH
theano
NVIDIA
jupyter
MAHOUT
MATLAB
R
APACHE Spark
MLlib

FROM RESEARCHERS TO PRACTITIONERS

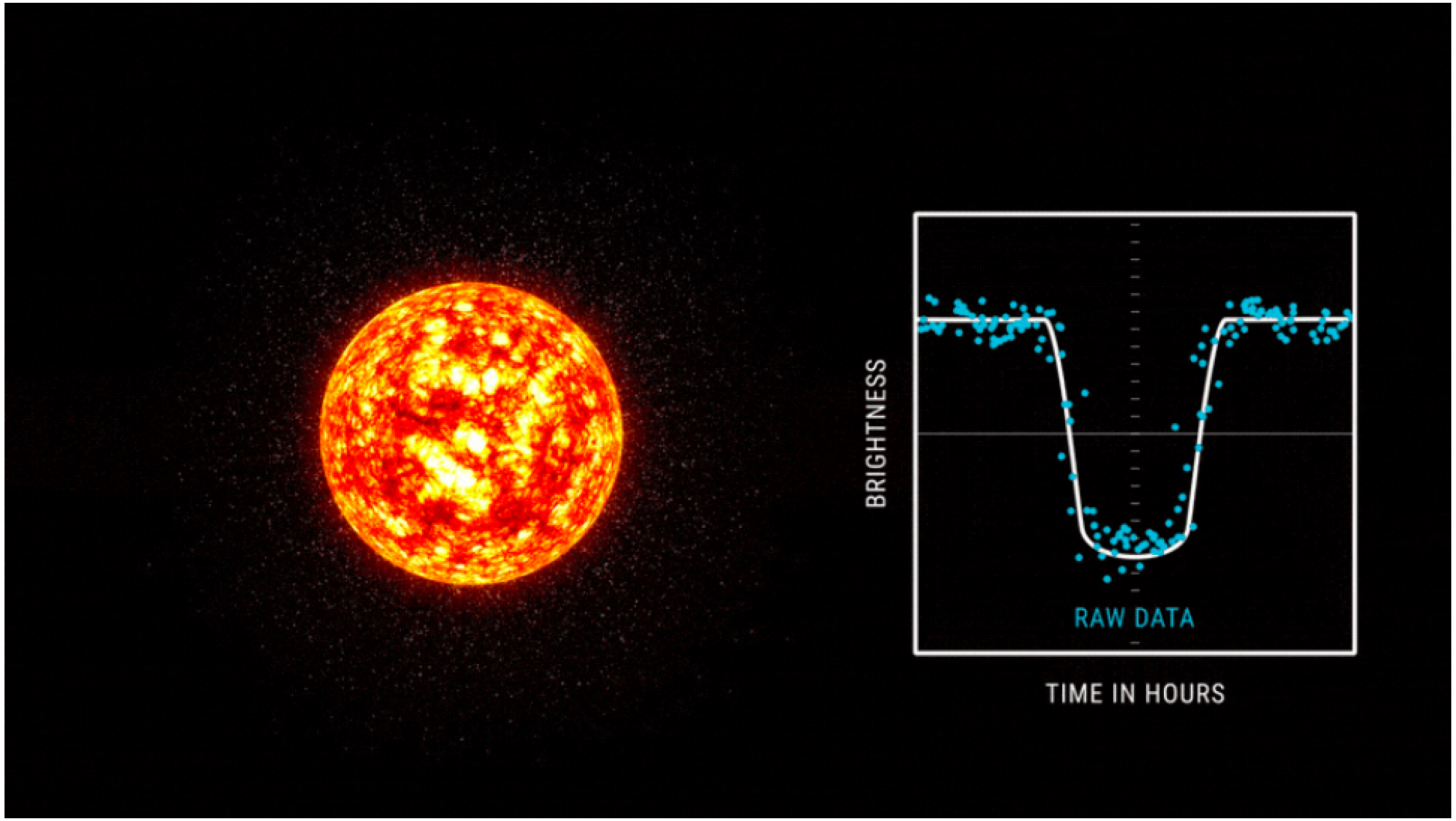
Gravitational Lensing



(10 Million Times Faster)

L. P. Levasseu

Hunting for Exoplanets (Kepler)



Chris Shallue

Assessing Cardiovascular Risk Factors with Computer Vision

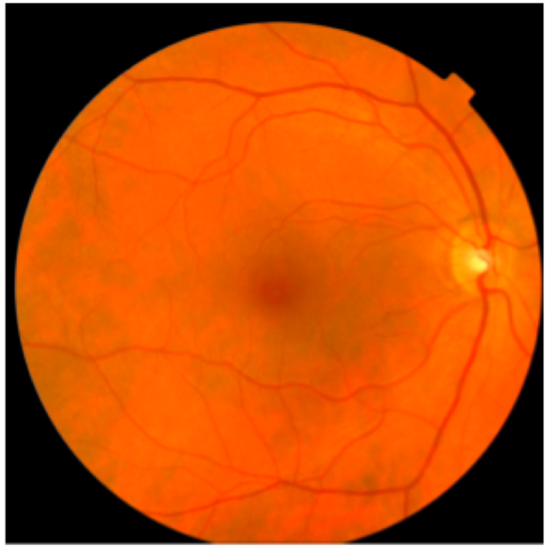
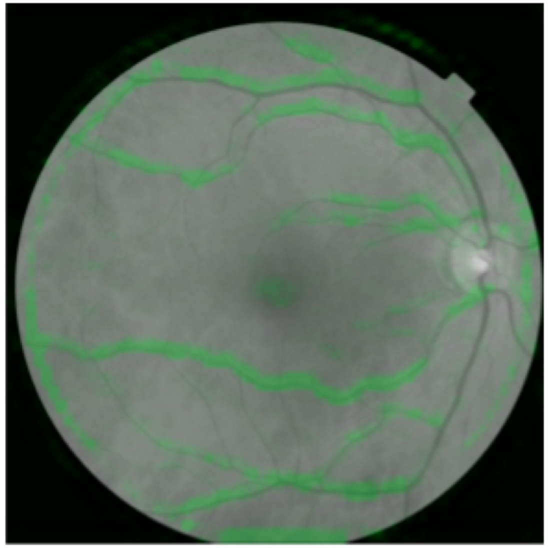


Image of retina

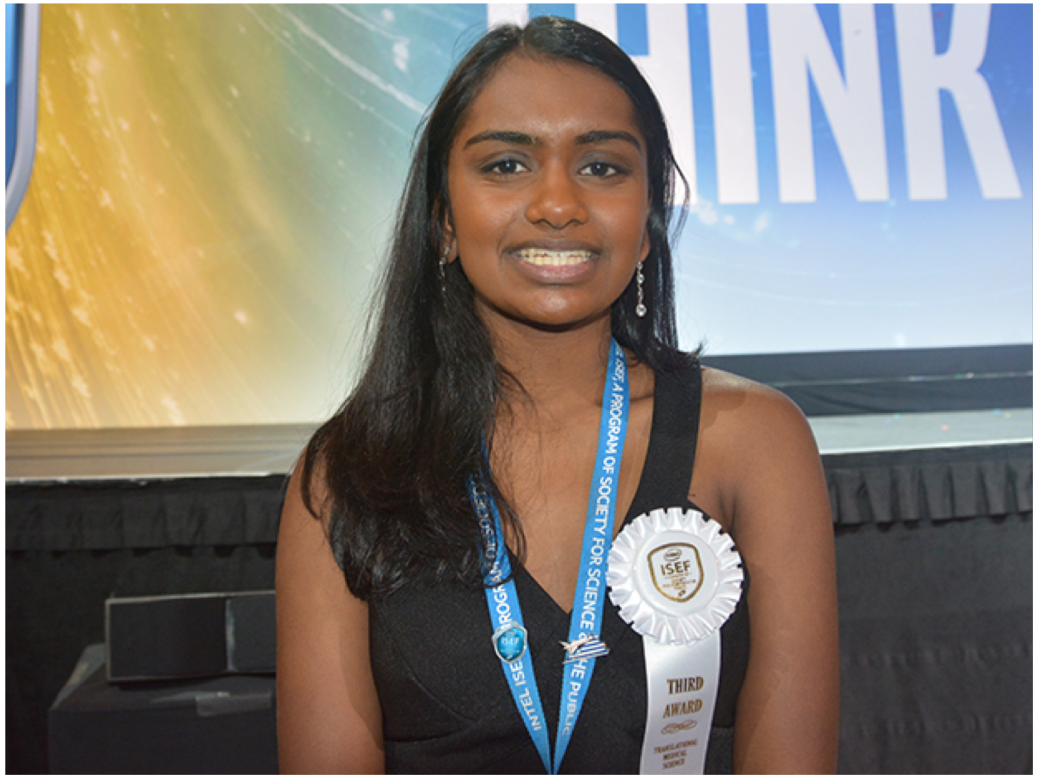


Blood pressure predictions focus on blood vessels

Lily Peng

High School Students Identifying Cancer

Abu Qadar



Kavya Kopparapu

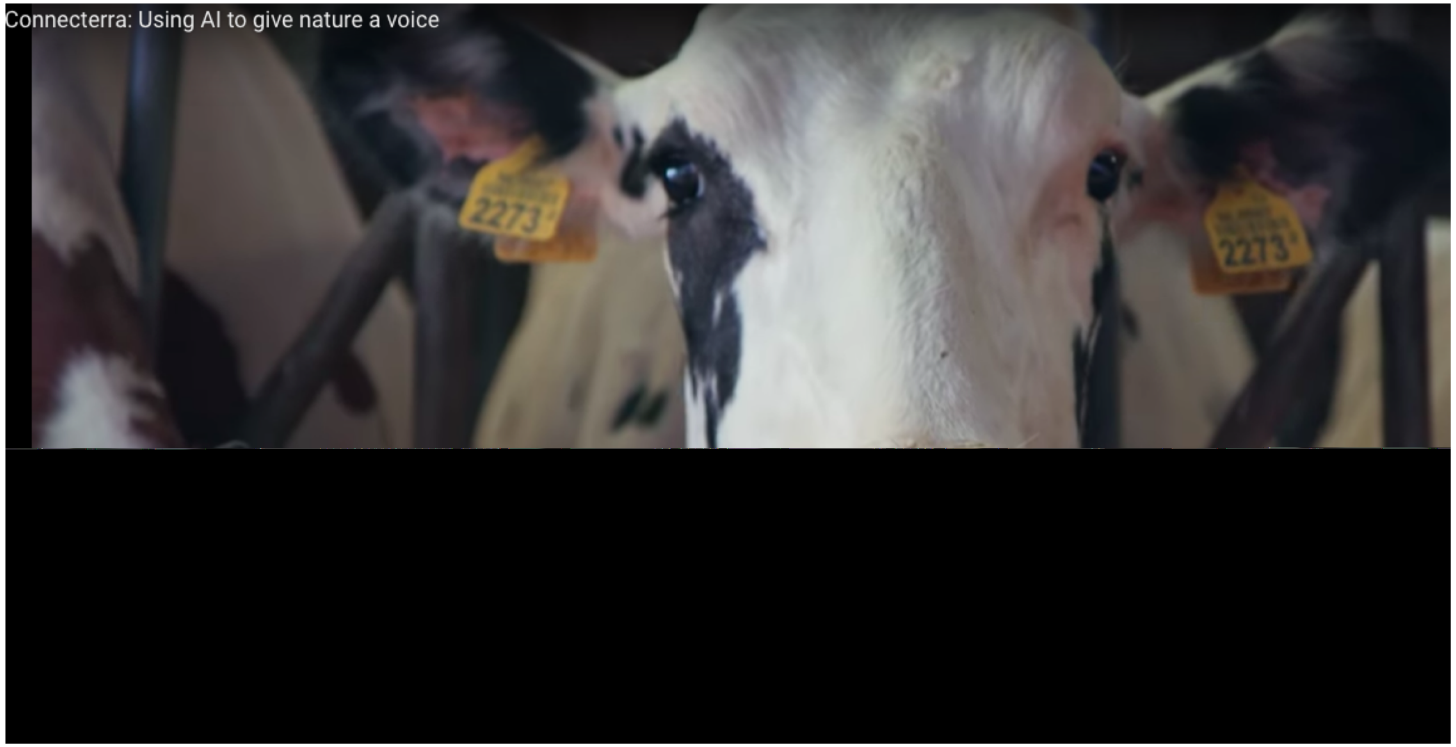
Identifying Sick Cassava Plants

Pete Warden

Fighting Illegal Deforestation

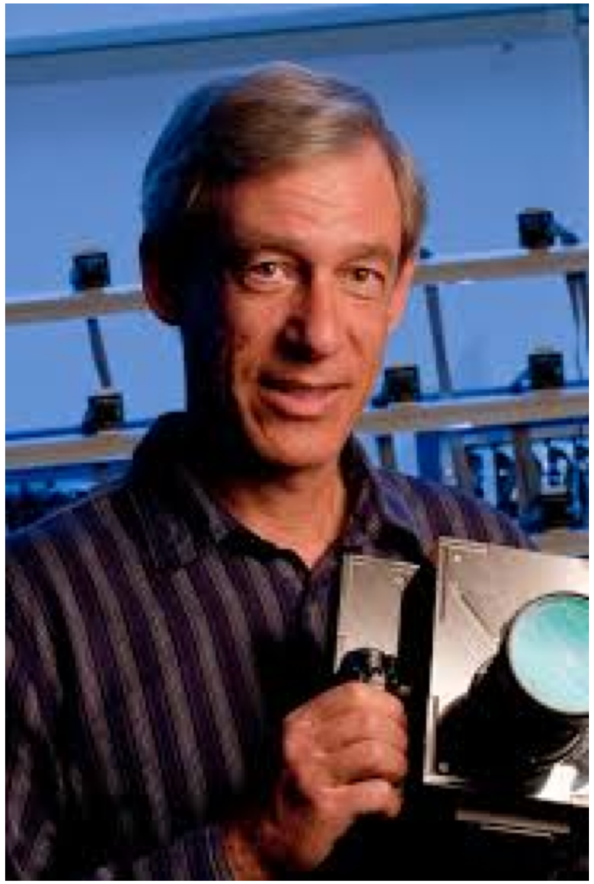
Topher White

Tracking Cows (Connecterra)



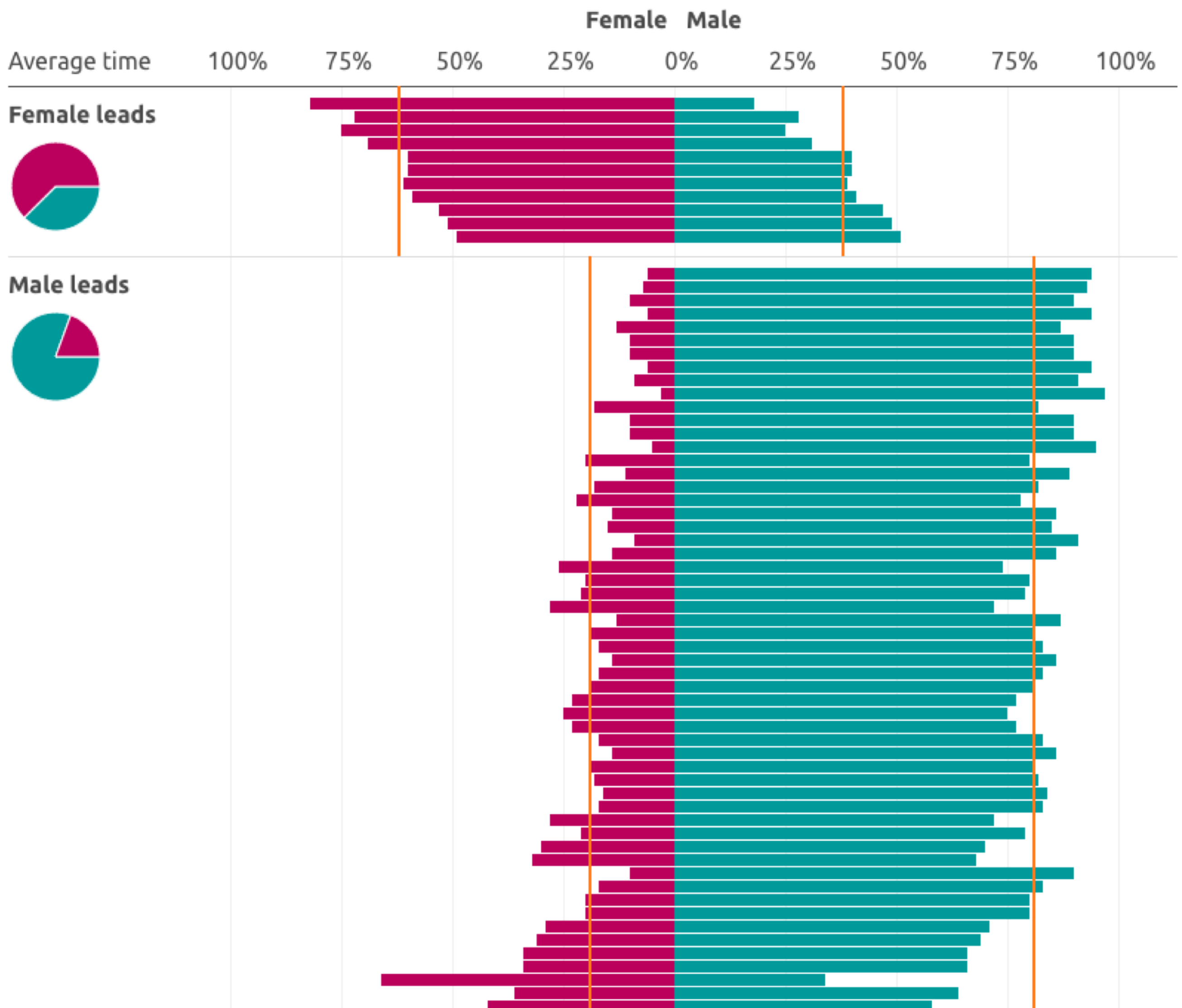
Saad Ansari

Portrait Mode



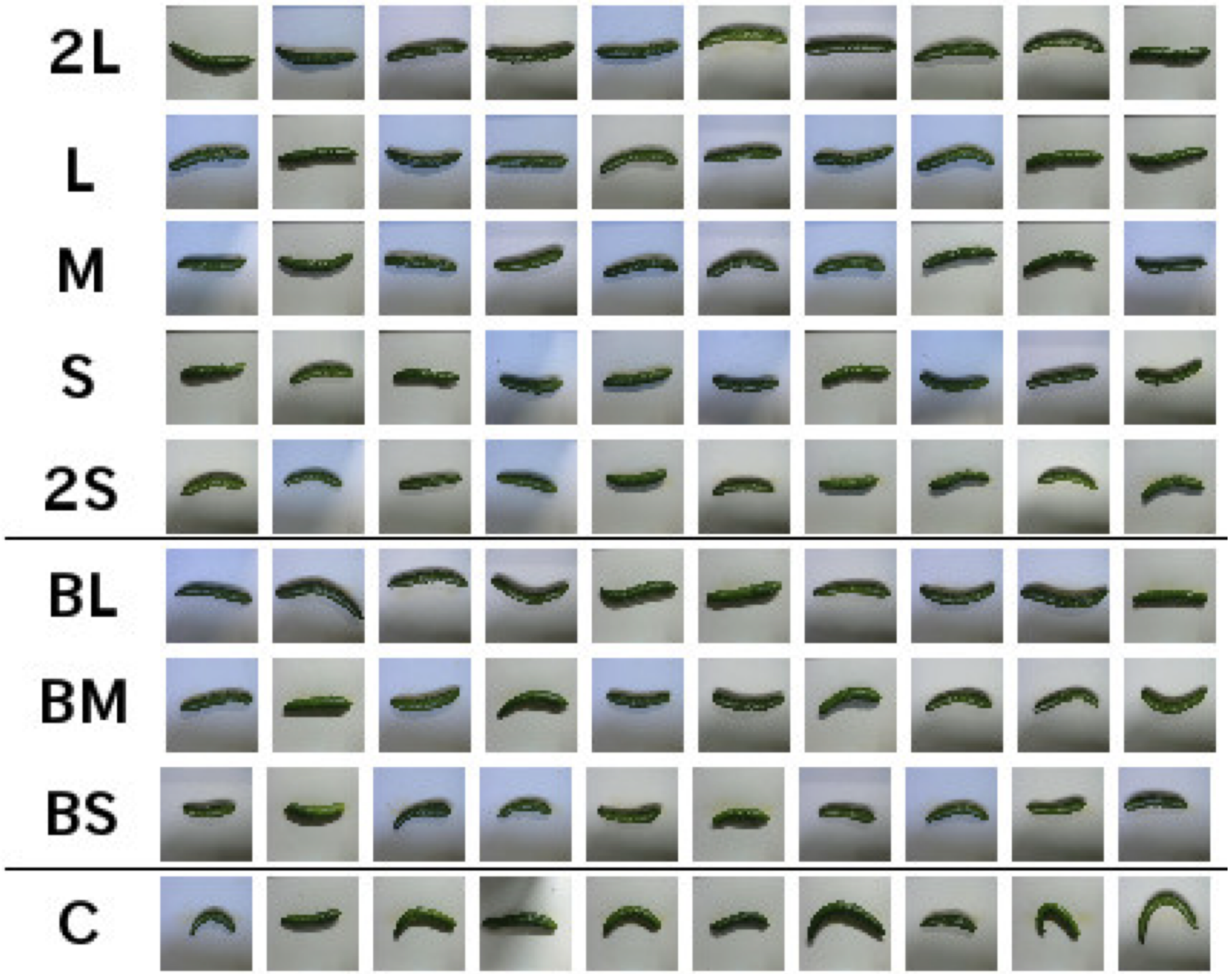
Marc Levoy

Uncovering Bias in Movies



Geena Davis

Sorting Cucumbers



Makoto Kolke

Things He Shouldn’t Have to Say

- There should be 17 convolution layers, then …
- The learning rate should be …
- Use L2 loss …
- To reduce training time, ignore these features …
- Use a onbe-hot encoding for this feature ,,,

Things He Should Say

- Here are examples of the nine classes
- The sorting station has electric lights, and also sunlight
- The worst mistake is to classify a C as something else.
- Look at size and thickness, color, texture, small scratches, whether or not they are crooked and whether they have prickles
- Prickles and scratches are less important
- This was classified wrong because …
- Sometimes if the camera is directly above the curve, it makes the curve hard to see
- The background doesn’t matter

Why ML?

Keeping Up with A Changing World

```
36.26.14 404 3 50 62
2012-10-15 00:38:23 100.43.227.77 GET /favicon.ico - 80 - 71.200.193.141 Mozilla/5.0+(Macintosh;+Intel+Mac+OS+X+10_7_5)+AppleWe
2012-10-15 00:38:23 100.43.227.77 GET /profiles/5248 - 80 - 68.48.125.234 Mozilla/5.0+(compatible;+MSIE+9.0;+Windows+NT+6.1;+WO
2012-10-15 00:38:26 100.43.227.77 GET /General-Contractor-Ephrata-PA - 80 - 66.249.73.218 Mozilla/5.0+(compatible;+Googlebot/2.
2012-10-15 00:39:37 100.43.227.77 GET /Baby-Accessories-in-ME - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+like
2012-10-15 00:39:38 100.43.227.77 GET /static/css/style.css v=2 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+like
2012-10-15 00:39:40 100.43.227.77 GET /Static/img/find-it.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+like
2012-10-15 00:39:40 100.43.227.77 GET /Static/img/relevant-yellow.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+
4
2012-10-15 00:39:40 100.43.227.77 GET /Static/img/btn_tweet.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+li
2012-10-15 00:39:40 100.43.227.77 GET /Static/img/btn_contact.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+
2012-10-15 00:39:40 100.43.227.77 GET /Static/img/printable-offer.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+
6
2012-10-15 00:39:40 100.43.227.77 GET /Static/img/ico-email.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+li
2012-10-15 00:39:40 100.43.227.77 GET /Static/img/coupon.jpg - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+like+
2012-10-15 00:39:41 100.43.227.77 GET /Static/img/send-message.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0
```

```
 GET /static/img/navbar-bg.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+li
 GET /Baby-Accessories-in-ME - 80 - 199.27.128.26 facebookexternalhit/1.1+(+http://www.faceboo
 GET /static/img/foot-bg.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+like
 GET /wp-content/themes/relevantyellow/type/museosans_900-webfont.woff - 80 - 108.162.216.25 M
8
 GET /wp-content/themes/relevantyellow/type/museosans_900-webfont.ttf - 80 - 108.162.216.23 Mo

 GET /wp-content/themes/relevantyellow/type/museosans_900-webfont.svg - 80 - 71.239.146.217 Mo
2
 GET /General-Contractor-Hamburg-NY - 80 - 66.249.73.227 Mozilla/5.0+(compatible;+Googlebot/2.
 GET /Beauty-Salon-East-Weymouth-MA/hours - 80 - 66.249.73.51 Mozilla/5.0+(compatible;+Googleb
 GET /Beauty-Salon-East-Weymouth-MA - 80 - 66.249.73.227 Mozilla/5.0+(compatible;+Googlebot/2.
 GET /static/img/foot-bg.jpg - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+like
 GET /Images/Logos/logo_70536.png - 80 - 66.249.73.51 Googlebot-Image/1.0 200 0 0 826
 GET /Mortgage-Lender-London-ON - 80 - 66.249.73.51 Mozilla/5.0+(compatible;+Googlebot/2.1;++h
 GET /static/img/claimed.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+like
 GET /static/img/searchbox-shadow.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS

 GET /static/img/nav-shadow.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+l
 GET /static/img/sidebar-bg.png - 80 - 71.239.146.217 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+6_0+l
 GET /Computer-Repairs-New-Hartford-NY - 80 - 174.252.41.99 Mozilla/5.0+(iPhone;+CPU+iPhone+OS

 GET /static/css/style.css v=2 80 - 174.252.41.99 Mozilla/5.0+(iPhone;+CPU+iPhone+OS+5_1_1+lik
 GET /Computer-Repairs-New-Hartford-NY - 80 - 199.27.128.81 facebookexternalhit/1.1+(+http://w
 GET /wp-content/themes/relevantyellow/type/museosans_900-webfont.woff - 80 - 173.245.52.246 M
2 78
 GET /wp-content/themes/relevantyellow/type/museosans_900-webfont.ttf - 80 - 173.245.52.245 Mo
 93
 GET /wp-content/themes/relevantyellow/type/museosans_900-webfont.svg - 80 - 174.252.41.99 Moz
 78
 GET /Concrete-Construction-Yuma-AZ - 80 - 157.55.35.47 Mozilla/5.0+(compatible;+bingbot/2.0;+
 GET /Real-Estate-Agent-Surrey-BC - 80 - 66.249.73.46 Mozilla/5.0+(compatible;+Googlebot/2.1;+
```

```
2012-10-15 00:39:42 100.43.227.77
2012-10-15 00:39:42 100.43.227.77
2012-10-15 00:39:42 100.43.227.77
2012-10-15 00:39:42 100.43.227.77
e/10A403+Safari/8536.25 404 0 2 7
2012-10-15 00:39:43 100.43.227.77
/10A403+Safari/8536.25 404 0 2 78
2012-10-15 00:39:46 100.43.227.77
/10A403+Safari/8536.25 404 3 50 6
2012-10-15 00:39:48 100.43.227.77
2012-10-15 00:39:48 100.43.227.77
2012-10-15 00:39:49 100.43.227.77
2012-10-15 00:40:18 100.43.227.77
2012-10-15 00:41:02 100.43.227.77
2012-10-15 00:41:09 100.43.227.77
2012-10-15 00:41:20 100.43.227.77
2012-10-15 00:41:20 100.43.227.77
8
2012-10-15 00:41:22 100.43.227.77
2012-10-15 00:41:22 100.43.227.77
2012-10-15 00:41:55 100.43.227.77
0 280
2012-10-15 00:41:56 100.43.227.77
2012-10-15 00:42:04 100.43.227.77
2012-10-15 00:42:05 100.43.227.77
ile/9B206+Safari/7534.48.3 404 0
2012-10-15 00:42:06 100.43.227.77
le/9B206+Safari/7534.48.3 404 0 2
2012-10-15 00:42:08 100.43.227.77
e/9B206+Safari/7534.48.3 404 3 50
2012-10-15 00:42:22 100.43.227.77
2012-10-15 00:42:30 100.43.227.77
```

Why ML?

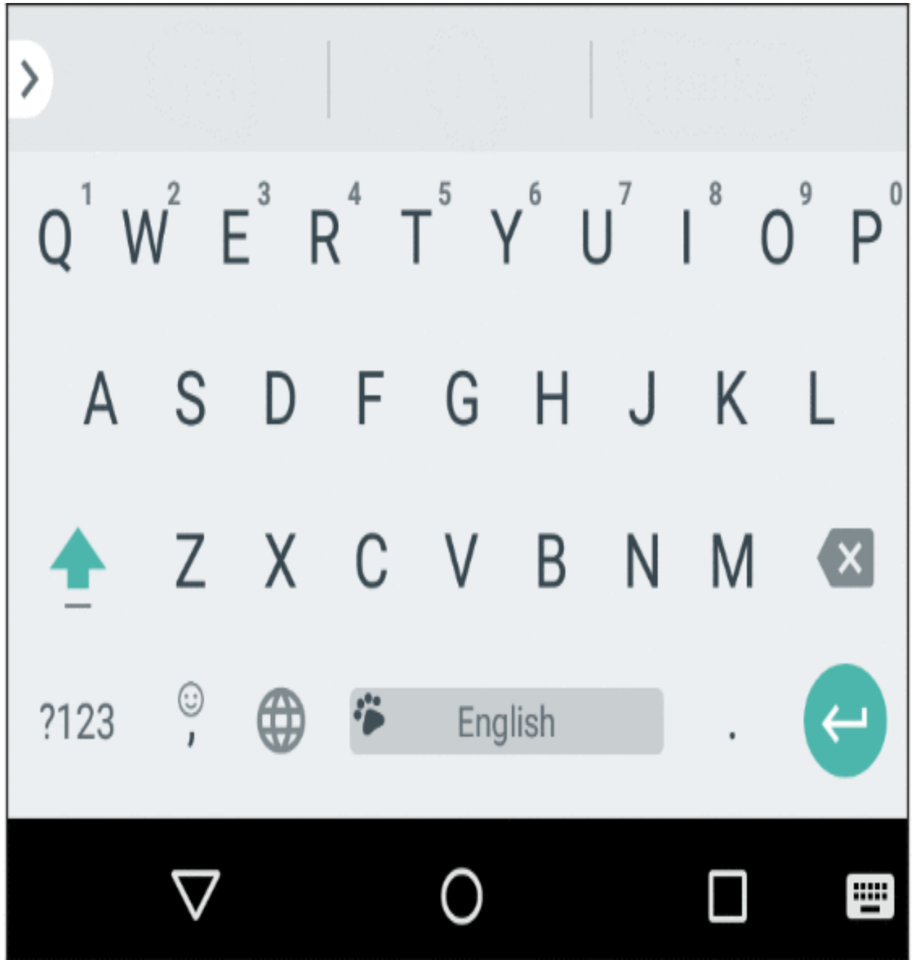
Development
Speed

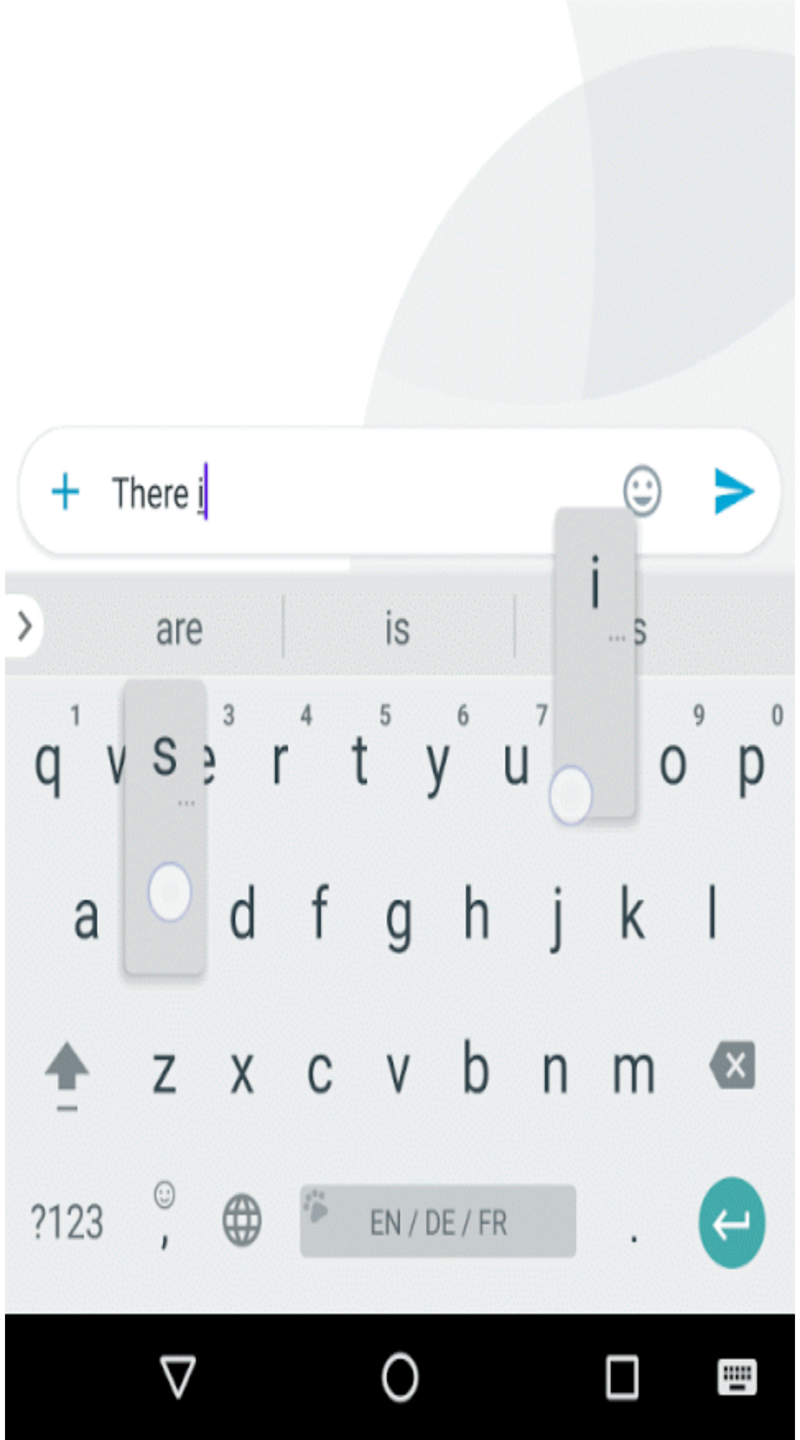

2
Safety Marshals
COASTLINE


32
TaxSlayer
TaxSlayer.com
88
CAMARO
BANDIT CHIPPERS
77
Just Imagine
99

Why ML?

Doing What You Don’t Know How to Do


Q W E R T Y U I O P
A S D F G H J K L
Z X C V B N M
?123
English


There i
are
is
q w e r t y u i o p
a s d f g h j k l
z x c v b n m
?123
EN / DE / FR

Dealing with Difficulties

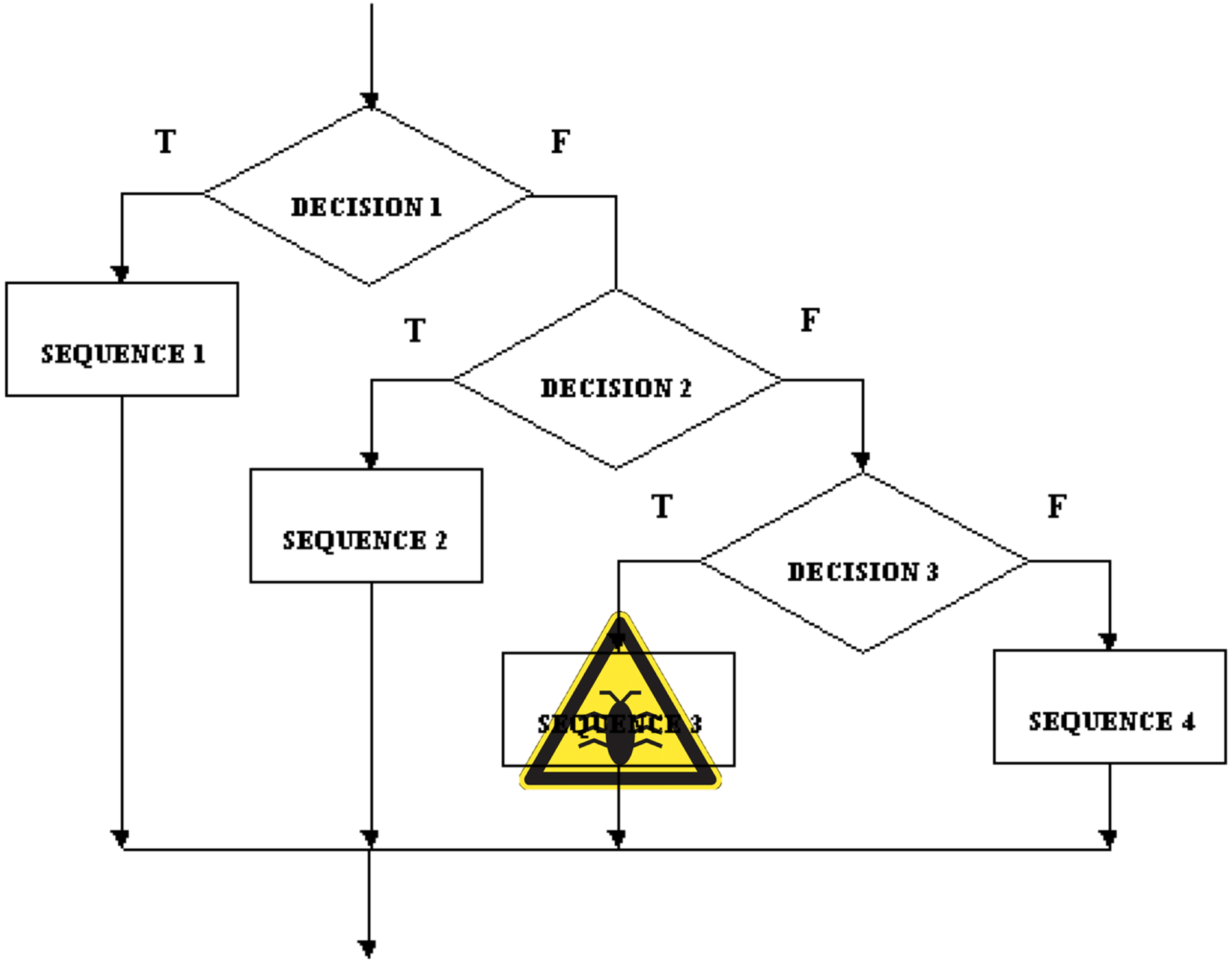
Machine Learning:
The High-Interest Credit Card of Technical Debt

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov,
Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young
{dsculley,gholt,dgg,edavydov}@google.com
{toddphillips,ebner,vchaudhary,mwyoung}@google.com
Google, Inc

Abstract

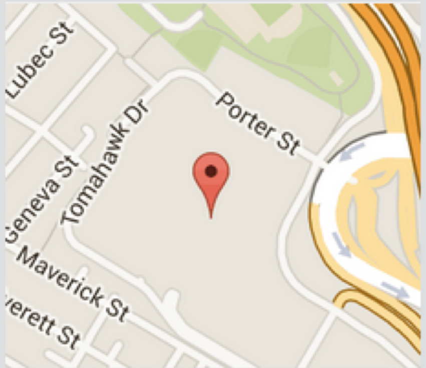
Machine learning offers a fantastically powerful toolkit for building complex systems quickly. This paper argues that it is dangerous to think of these quick wins as coming for free. Using the framework of *technical debt*, we note that it is remarkably easy to incur massive ongoing maintenance costs at the system level when applying machine learning. The goal of this paper is highlight several machine learning specific risk factors and design patterns to be avoided or refactored where possible. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, changes in the external world, and a variety of system-level anti-patterns.

Lack of Clear Abstraction Barriers



Example: Google Now

Depart now for:
Return Rental Car to Logan Airport
156 Tomahawk Dr,
Boston MA

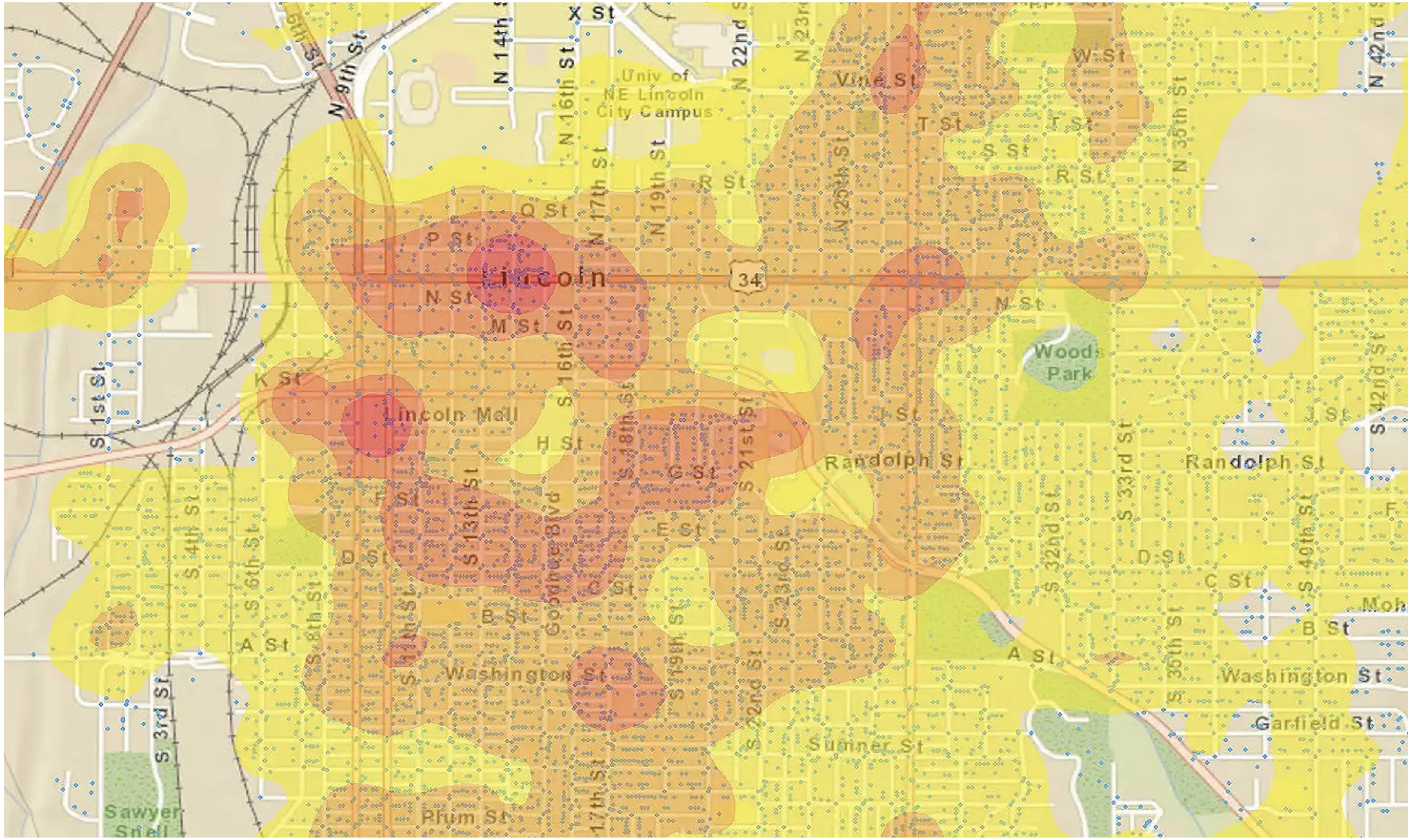


Time of travel:
23 minutes by bicycle

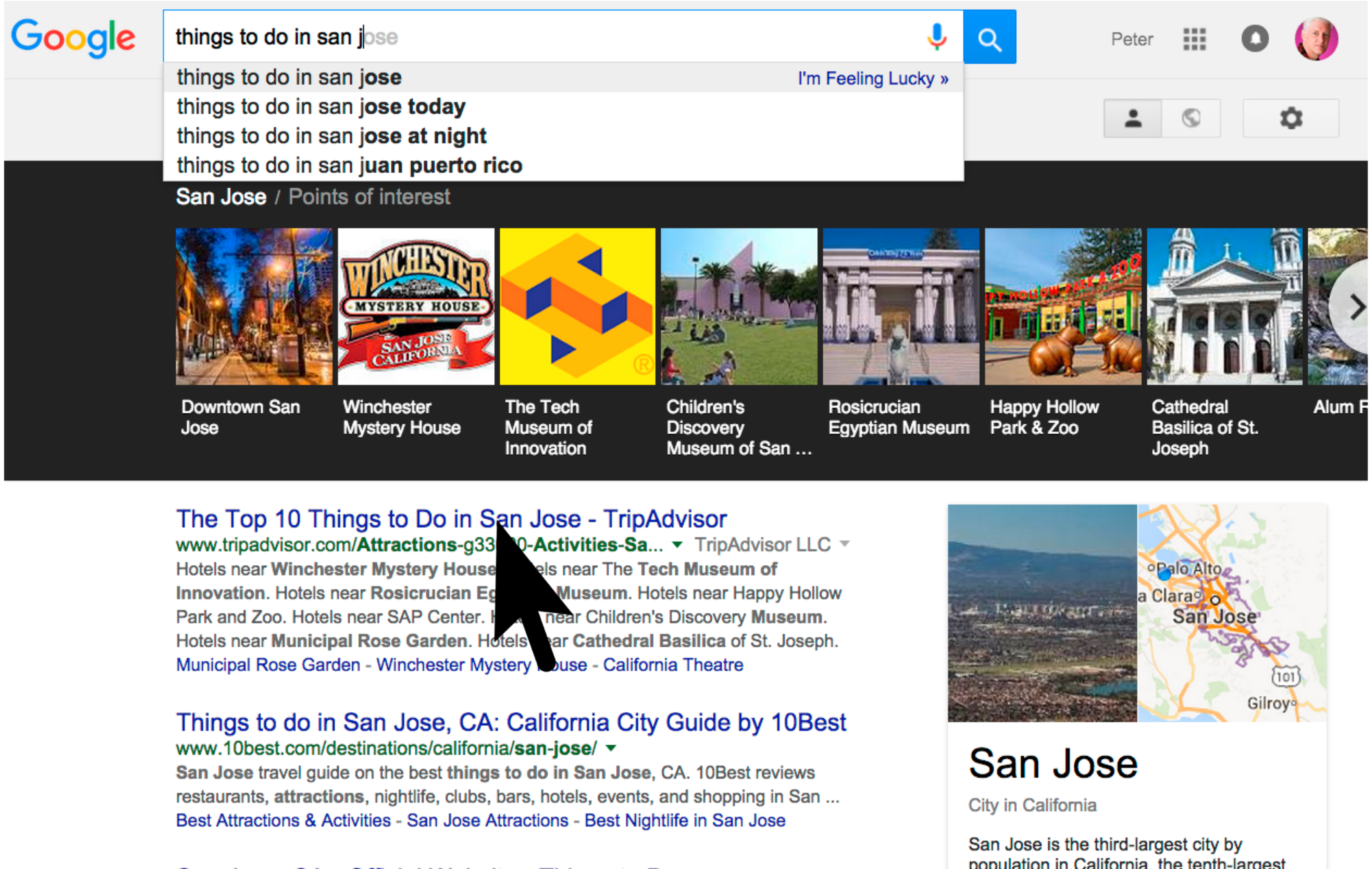
```
event = ExtractEvent(email.body)
trip  = Travel(current.location, event.location, event.time)
CreateAlert(trip)
```

Nonstationarity

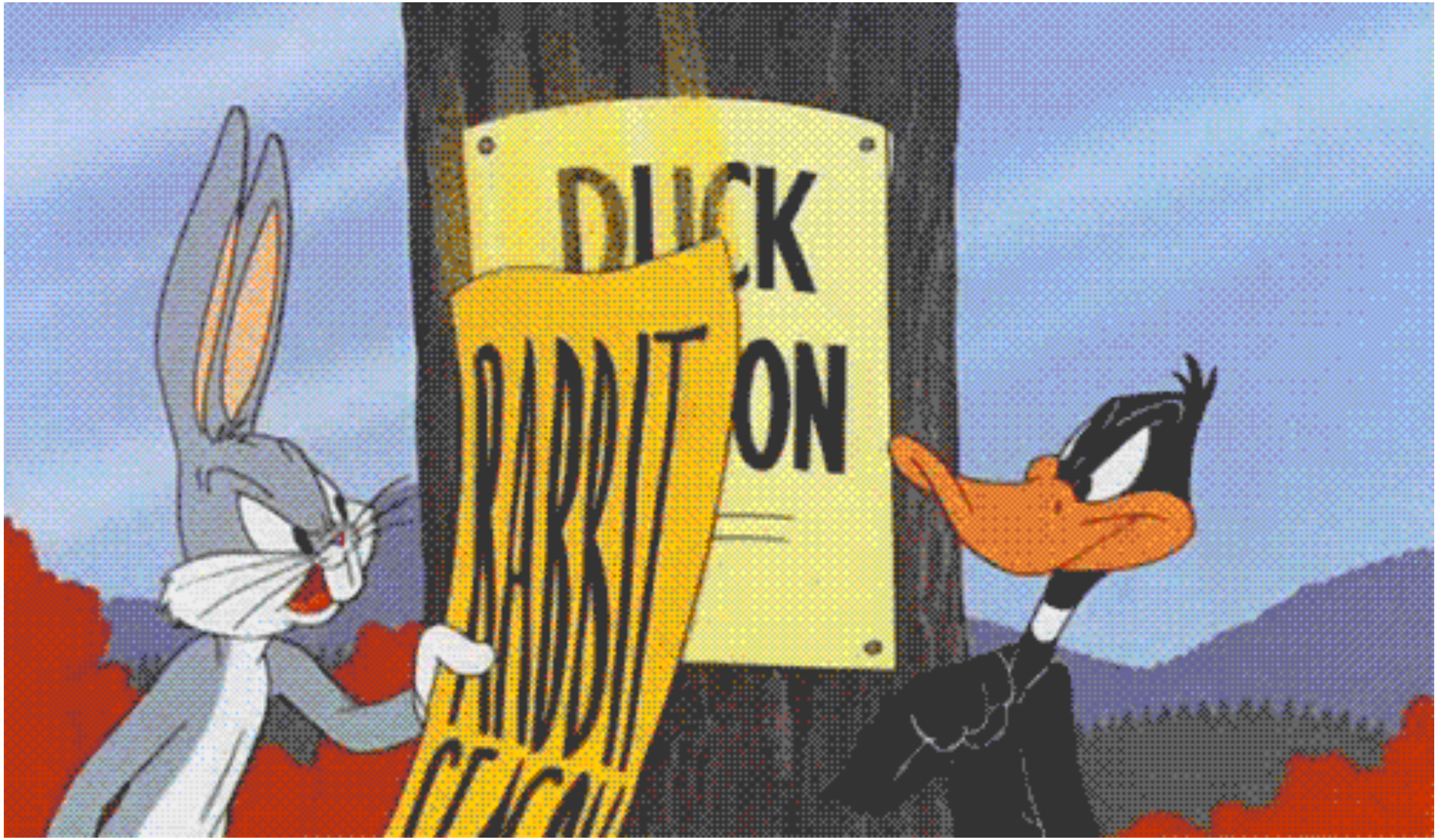
Feedback Loops



Feedback Loops

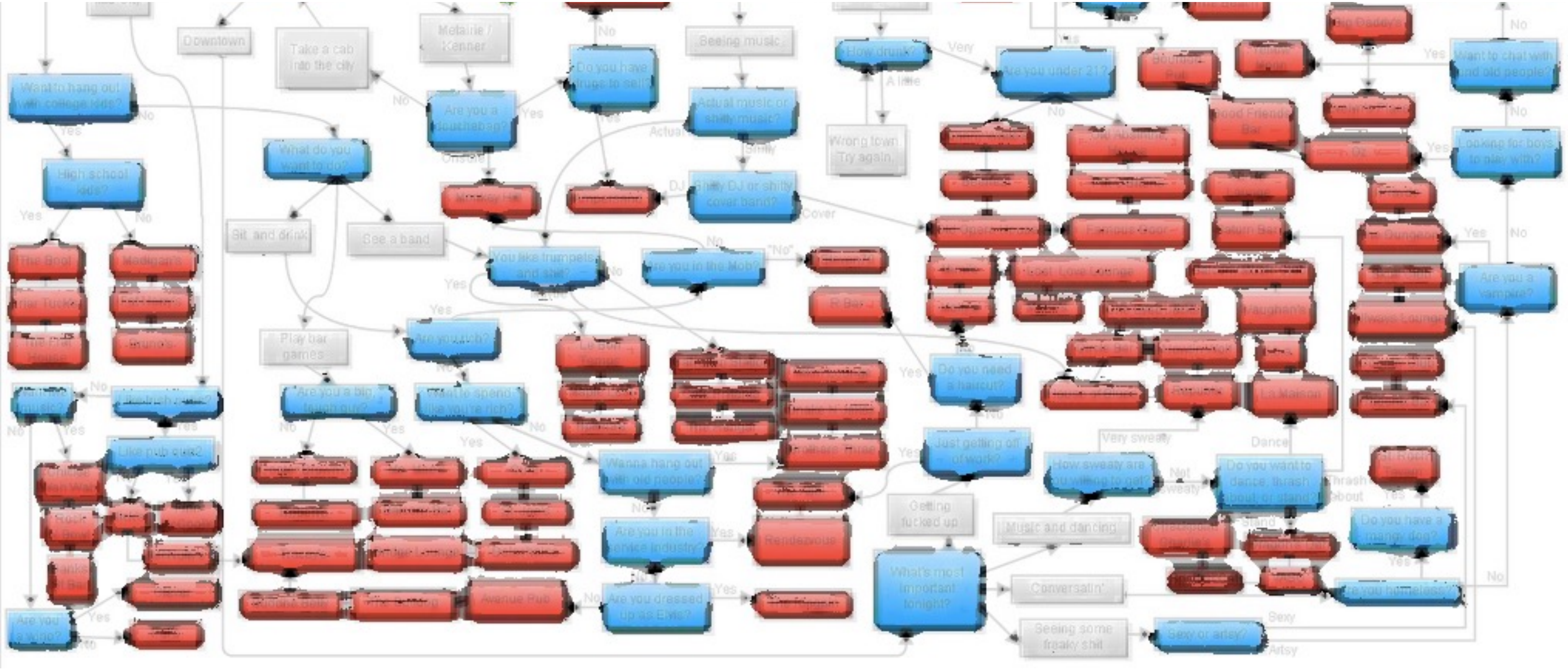


Feedback Loops



Privacy, Security, Fairness

Data/Configuration Dependencies



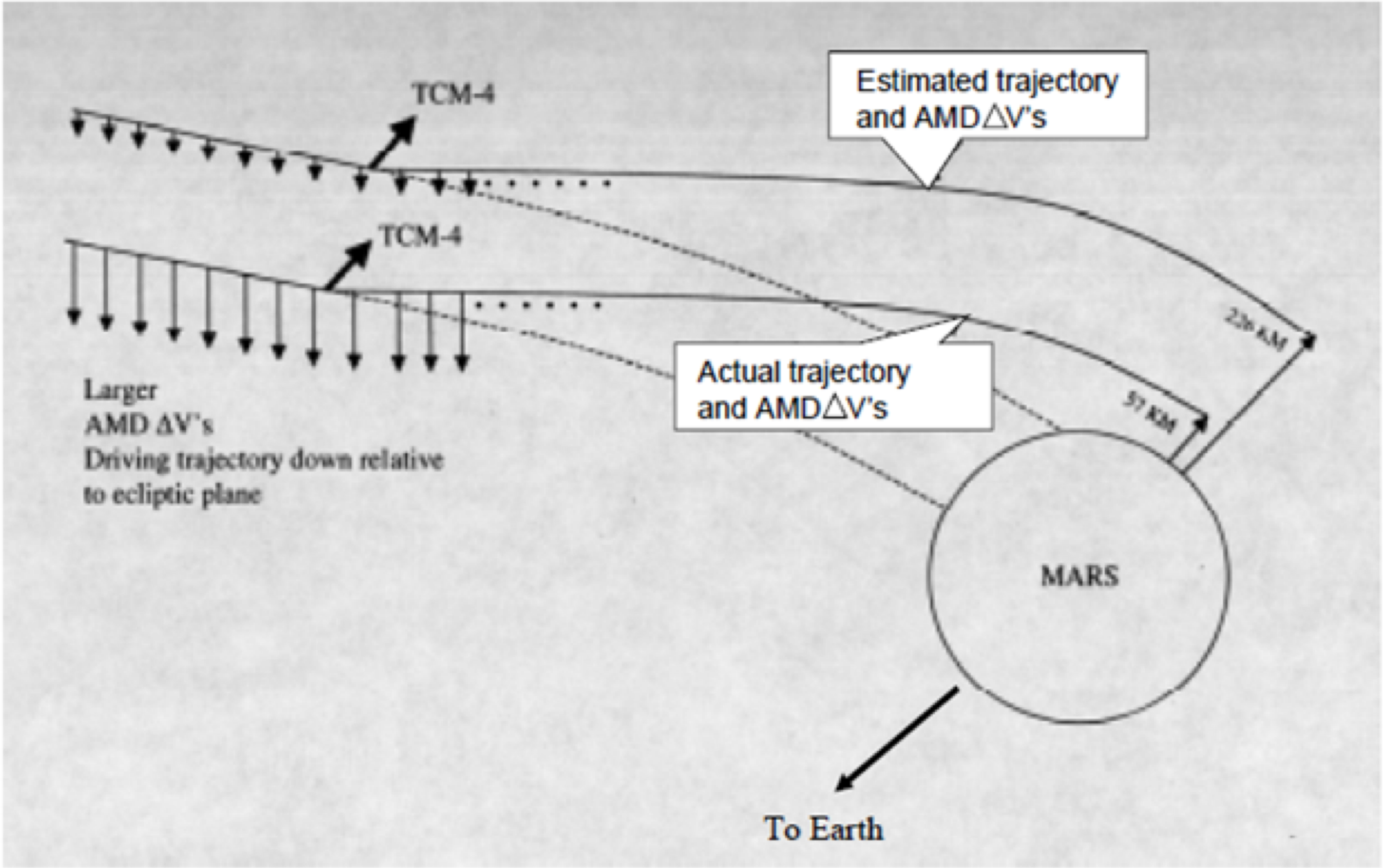
Data Dependencies:
Blacklist: do not index

really-gross-stuff.com
calendar/next-page-button
lots-of-other/things

/cgi/
/cgi-bin/

Schematic MCO Encounter Diagram

Not to scale



Concrete Problems in AI Safety

Dario Amodei* Google Brain

Chris Olah* Google Brain

Jacob Steinhardt Stanford University

Paul Christiano UC Berkeley

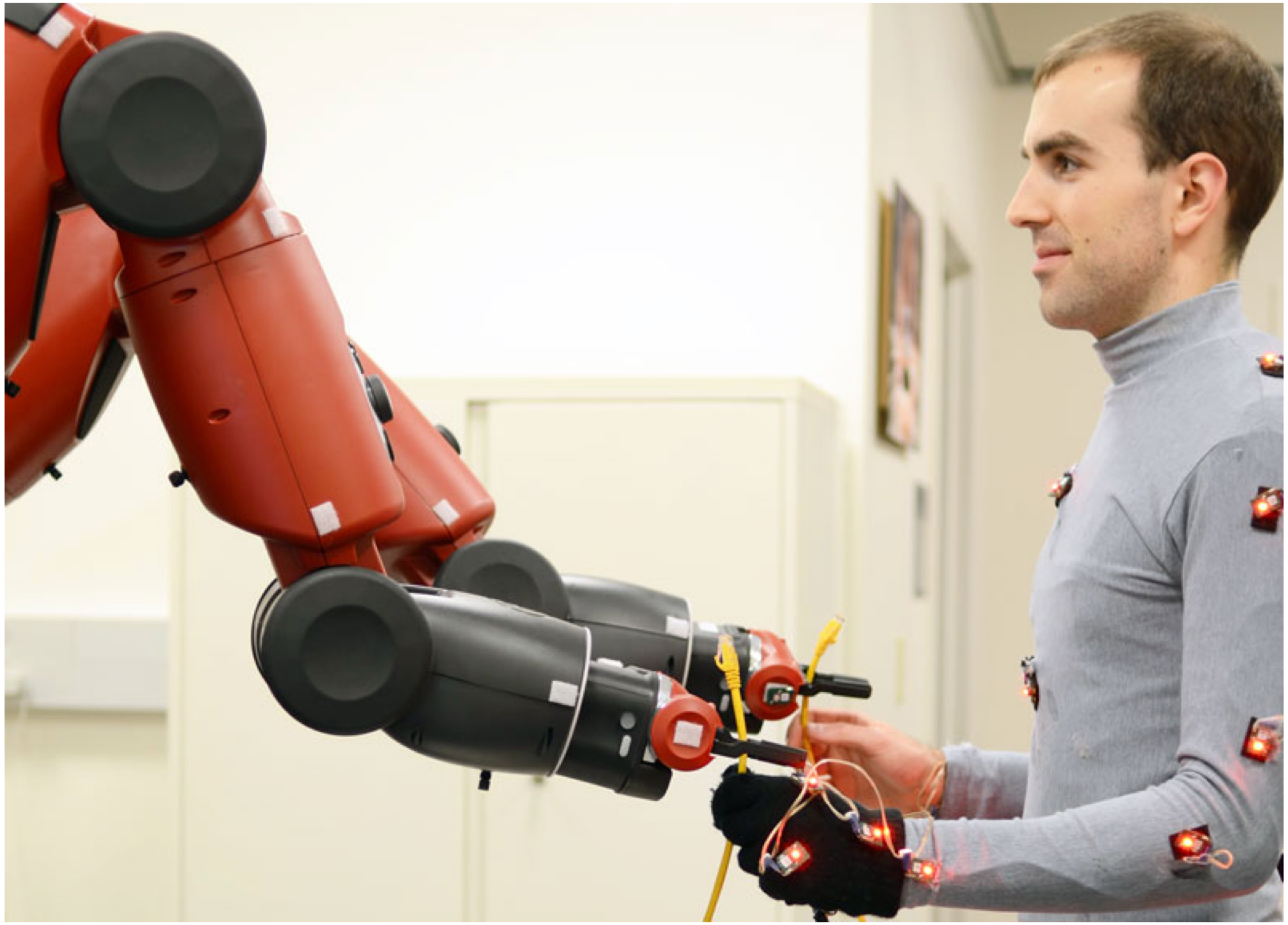
John Schulman OpenAI

Dan Mané Google Brain

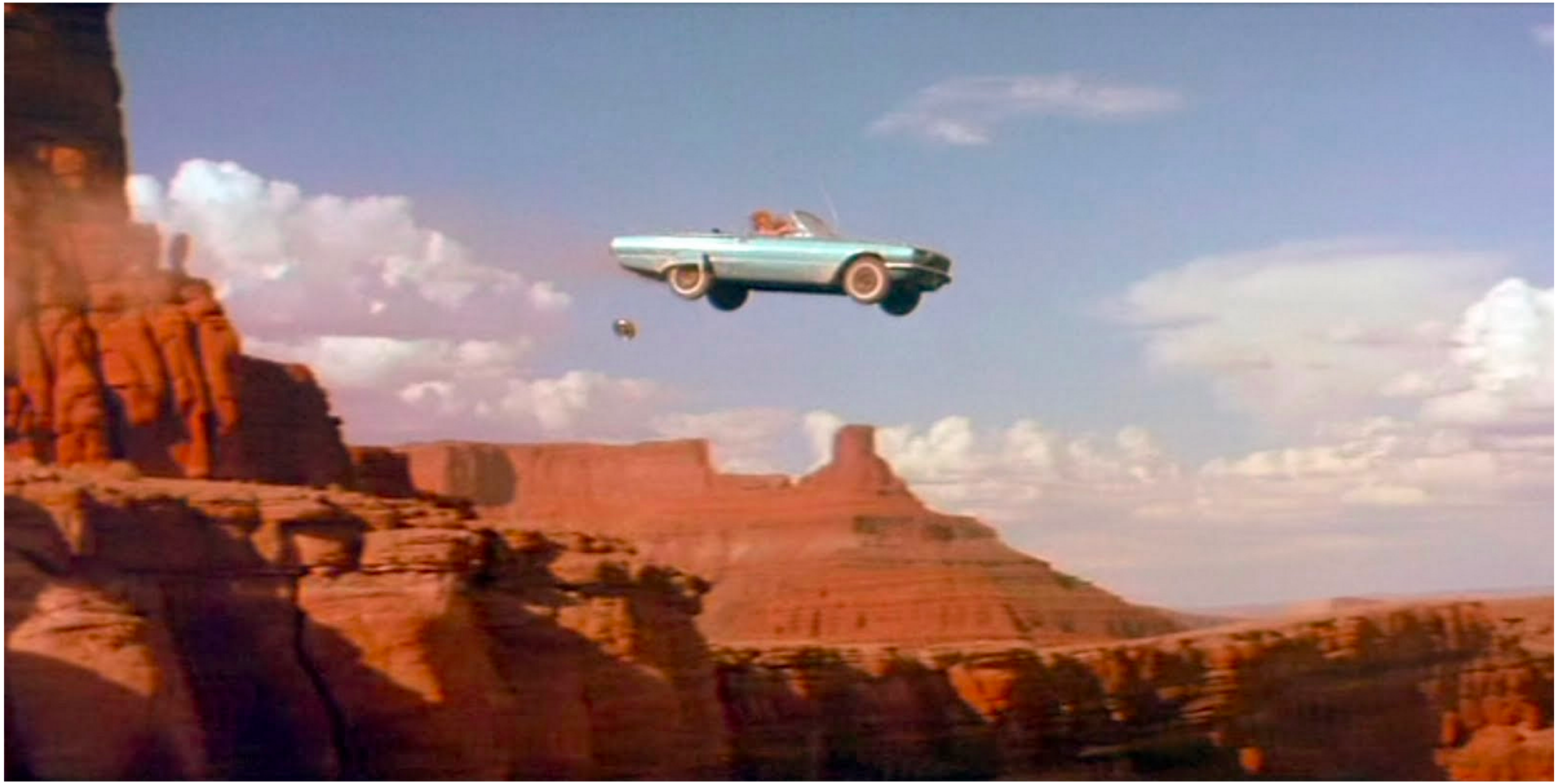
Abstract

Rapid progress in machine learning and artificial intelligence (AI) has brought increasing attention to the potential impacts of AI technologies on society. In this paper we discuss one such potential impact: the problem of *accidents* in machine learning systems, defined as unintended and harmful behavior that may emerge from poor design of real-world AI systems. We present a list of five practical research problems related to accident risk, categorized according to whether the problem originates from having the wrong objective function ("avoiding side effects" and "avoiding reward hacking"), an objective function that is too expensive to evaluate frequently ("scalable supervision"), or undesirable behavior during the learning process ("safe exploration" and "distributional shift"). We review previous work in these areas as well as suggesting research directions with a focus on relevance to cutting-edge AI systems. Finally, we consider the high-level question of how to think most productively about the safety of forward-looking applications of AI.

Scalable Oversight



Safe Exploration



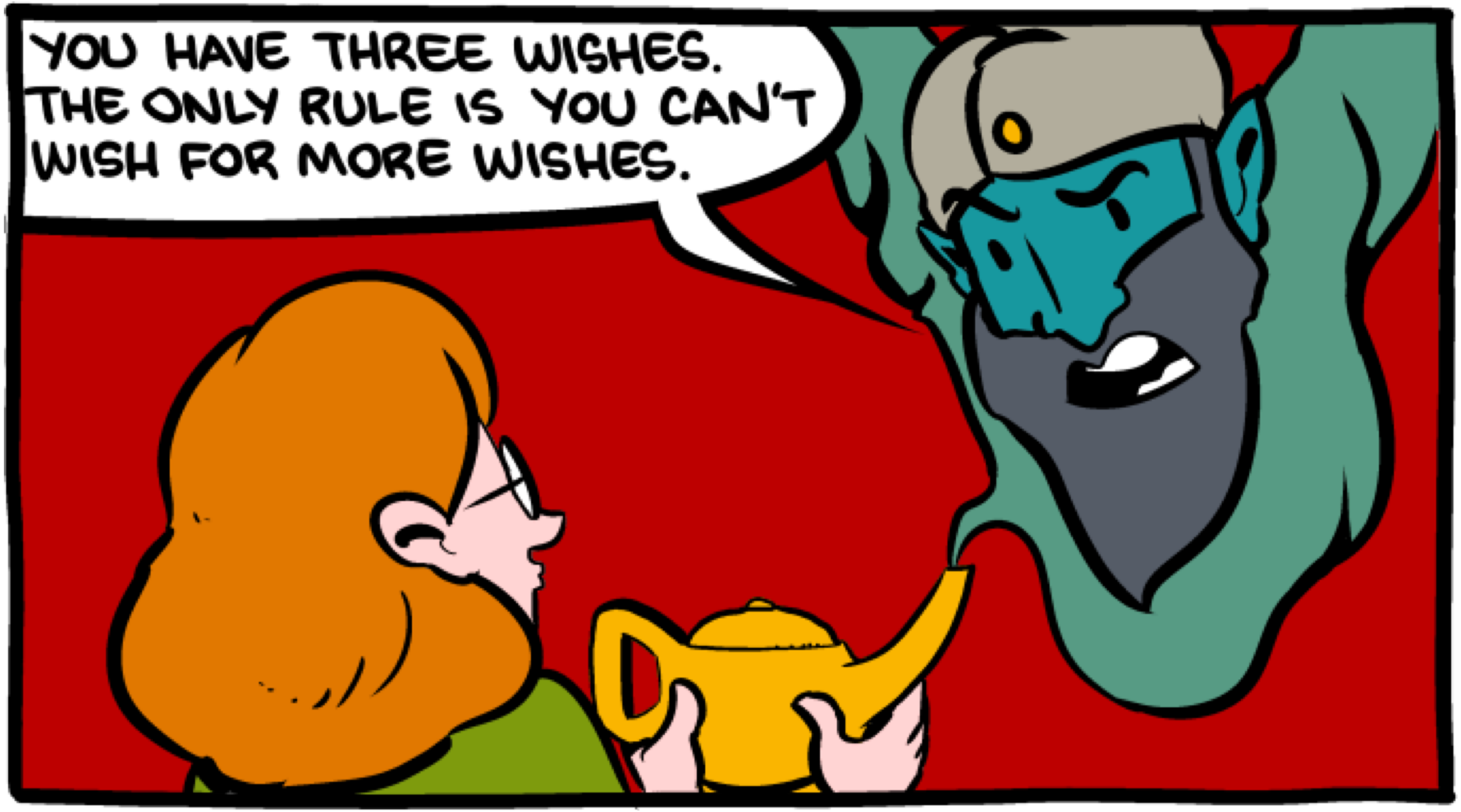
Inattention Valley

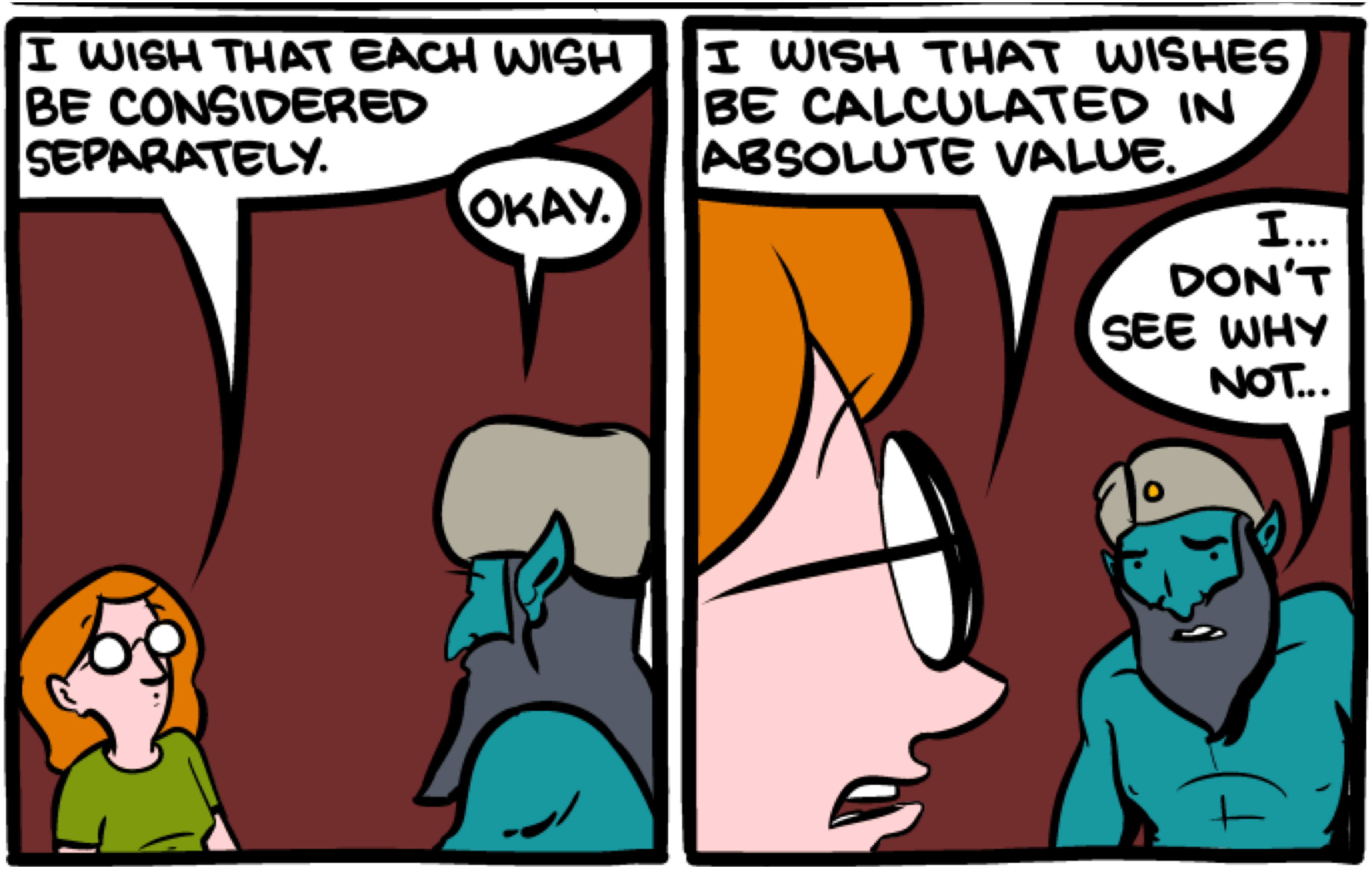
Utility Function Design

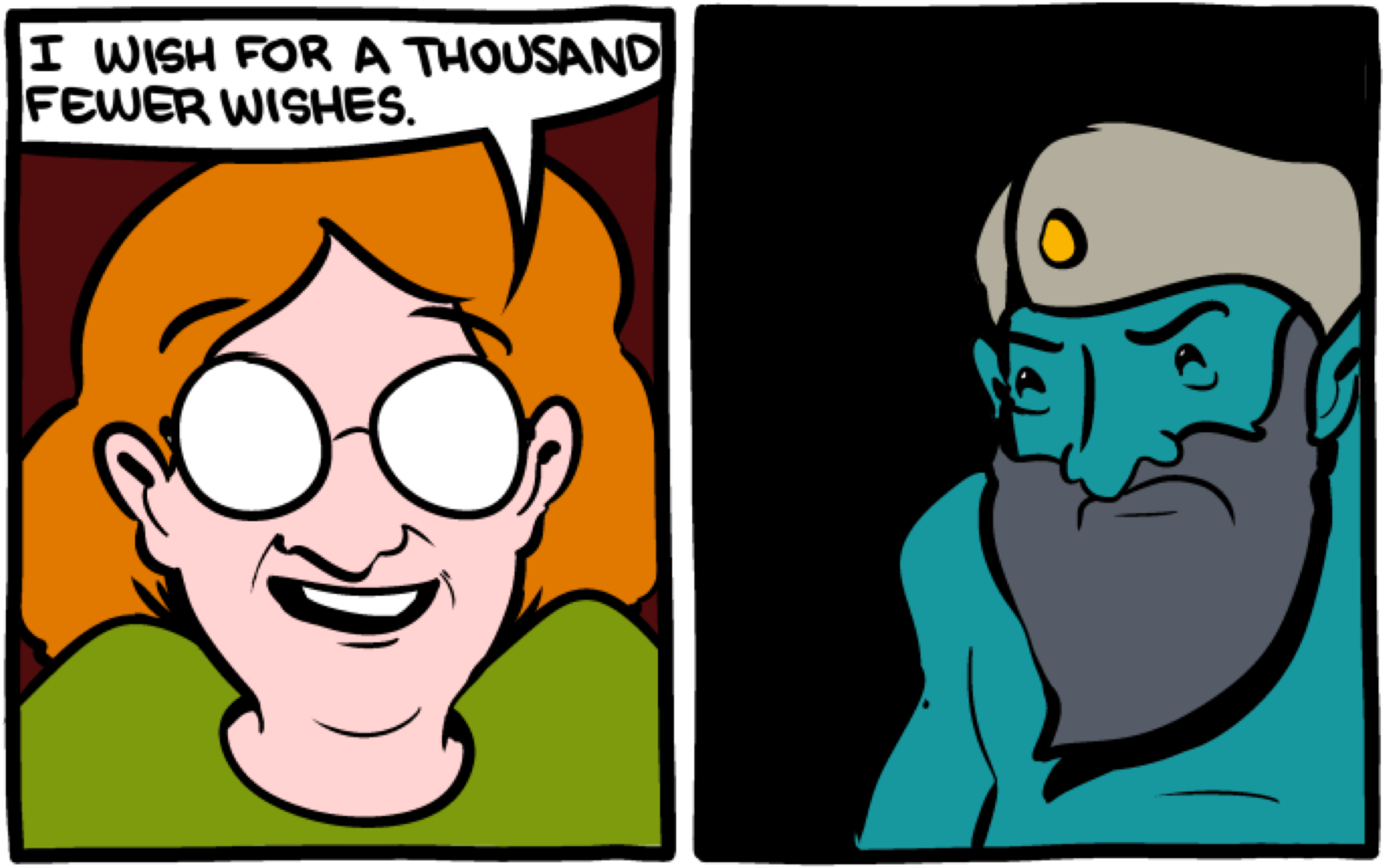
Fundamental Formula of AI

$$\text{act}^* = \text{argmax}_{\text{a in Actions}}\ E(\text{Utility}(\text{Result}(a, s)))$$

- State Estimation: s
- Model of World: Actions, Result
- Probabilistic Reasoning: E
- Search Algorithm: argmax
- Our Values/Desires: Utility


YOU HAVE THREE WISHES. THE ONLY RULE IS YOU CAN'T WISH FOR MORE WISHES.


I WISH THAT EACH WISH BE CONSIDERED SEPARATELY.
OKAY.
I WISH THAT WISHES BE CALCULATED IN ABSOLUTE VALUE.
I... DON'T SEE WHY NOT...

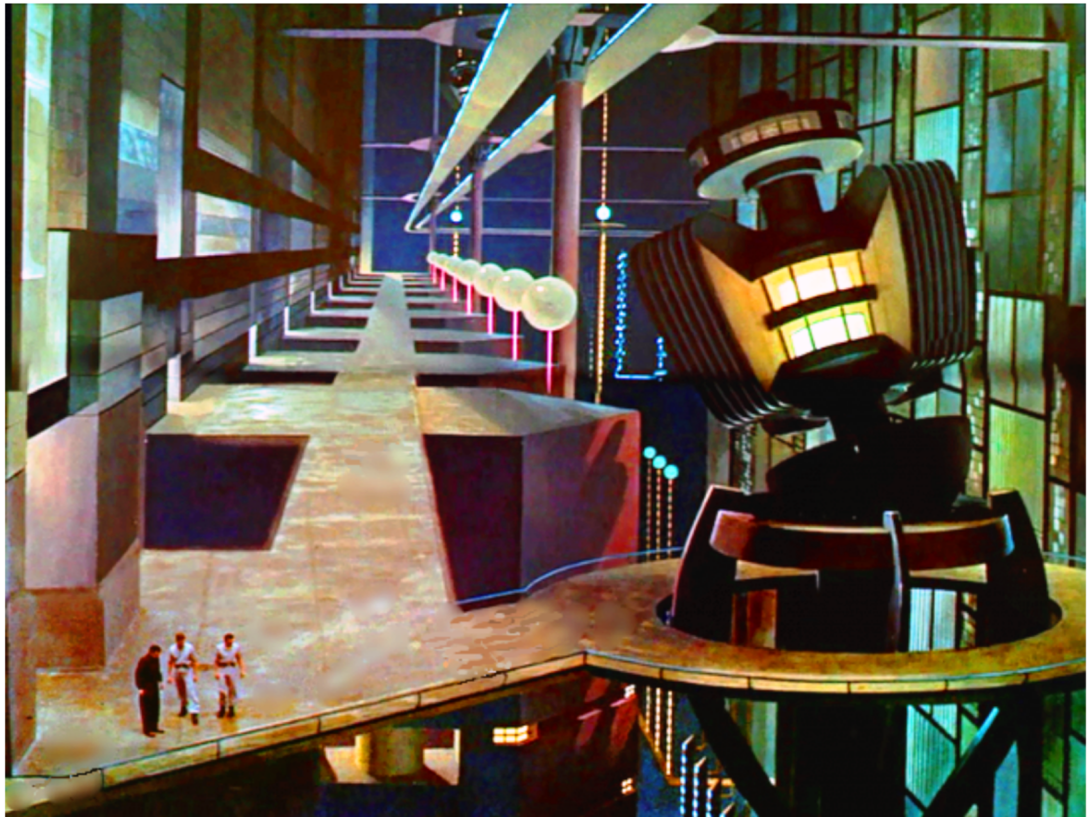

I WISH FOR A THOUSAND FEWER WISHES.


THIS IS WHY NO ONE LIKES MATHEMATICIANS.
THIS IS *ONE* REASON WHY.

Last Tweets of the Krell

Peter Norvig Follow
Mar 31, 2017 · 5 min read

Many readers are no doubt familiar with *Forbidden Planet*, the documentary film about the *Krell* civilization, which came to an unfortunate end just at the launch of what could have been their biggest achievement. Ever since the film's release in 1956, xenoanthropologists have been stymied by a lack of source material on the Krell.



Bug Fixes for Midas / Genie / Krell

- Specify an "undo" function
- Do some simulations first
- Then do small-scale tests
- Do adversarial tests / red teams
- Monitor tests and deployments
- Have ability to shut down or roll back
- Specify a prompt for unsafe actions
- Continuous evolution and improvement
- Don't over-rely on language: case law

WHAT WILL PROGRAMMERS BE?

PROGRAMMER = ~~MICROMANAGER?~~
TEACHER?   GENERAL?   PHILOSOPHER?

Don’t tell people **how** to do things, tell them **what** to do, and let them surprise you with their results.

– Gen. George Patton

SCIENTIST!