



# Testing AI-Based Systems

Jason Arbon





Login

- Startups
- Apps
- Gadgets
- Videos
- Podcasts
- Extra Crunch
- 
- Events
- Advertise
- Crunchbase
- More

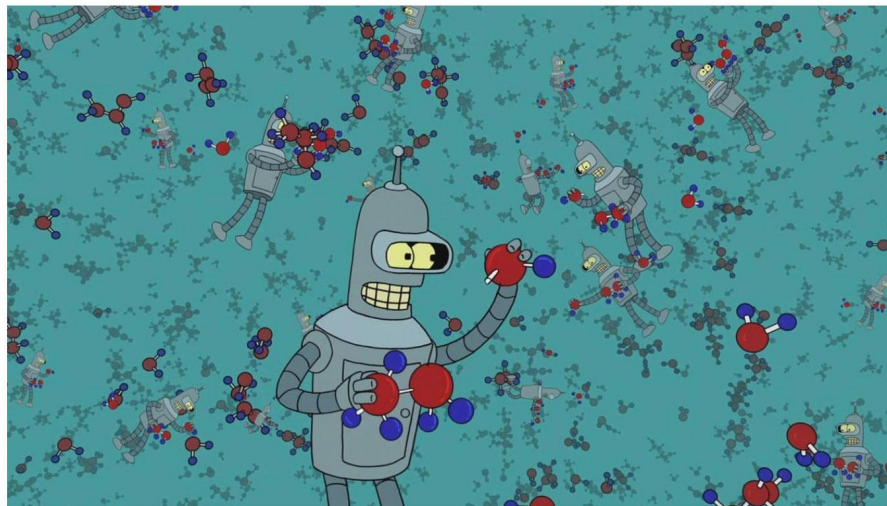
- Cybersecurity 101
- Fintech
- Transportation
- Google

# Test.ai nabs \$11M Series A led by Google to put bots to work testing apps



Lucas Matney @lucasmtny / 7 months ago

Comment



For developers, the process of determining whether every new update is going to botch some core functionality can take up a lot of time and resources, and things get far more complicated when you're managing a multitude of apps.

test.ai

# We're building the future of automated testing

Stay tuned for updates

Get on the List



## Overview

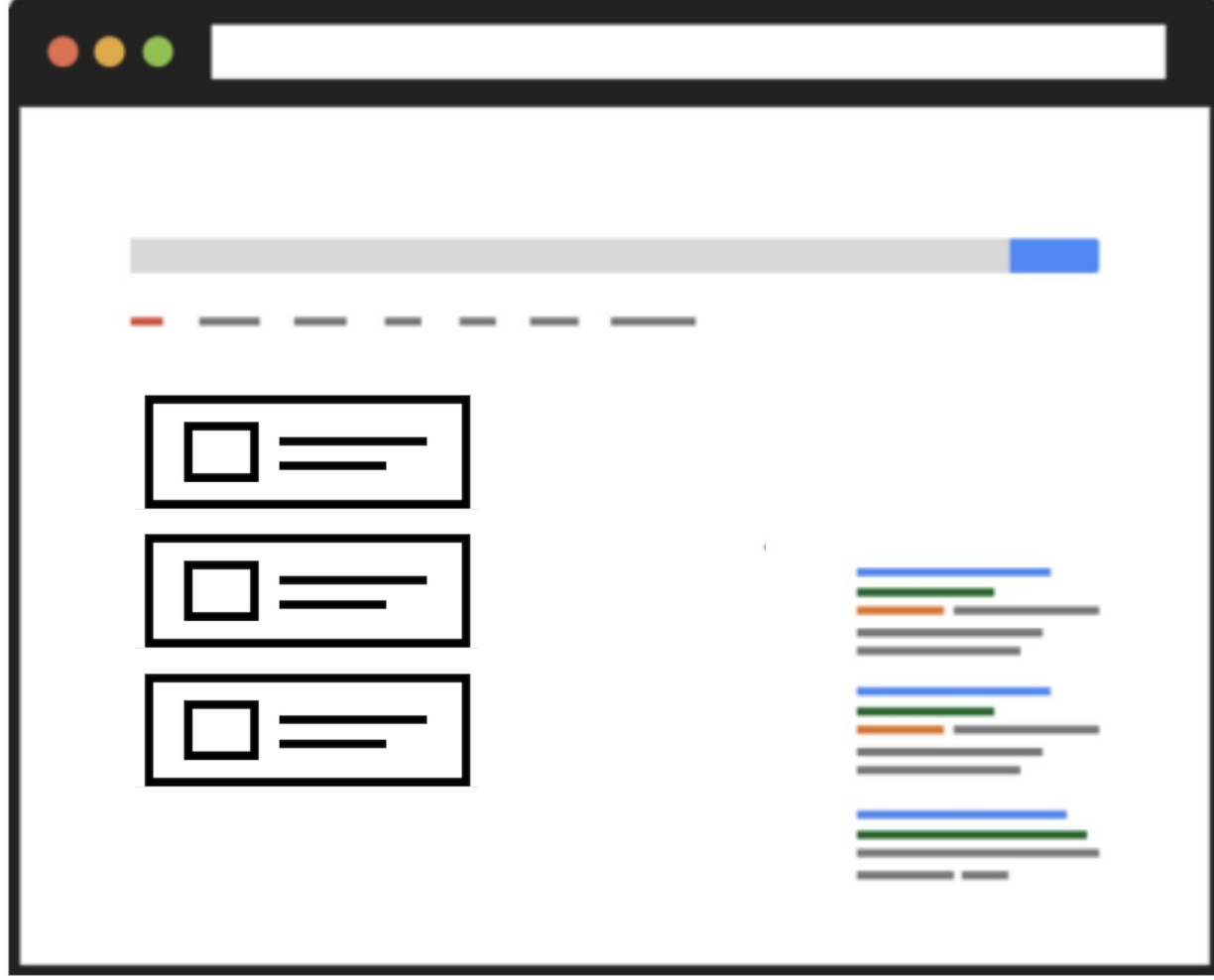
Test Training Data

Test Training

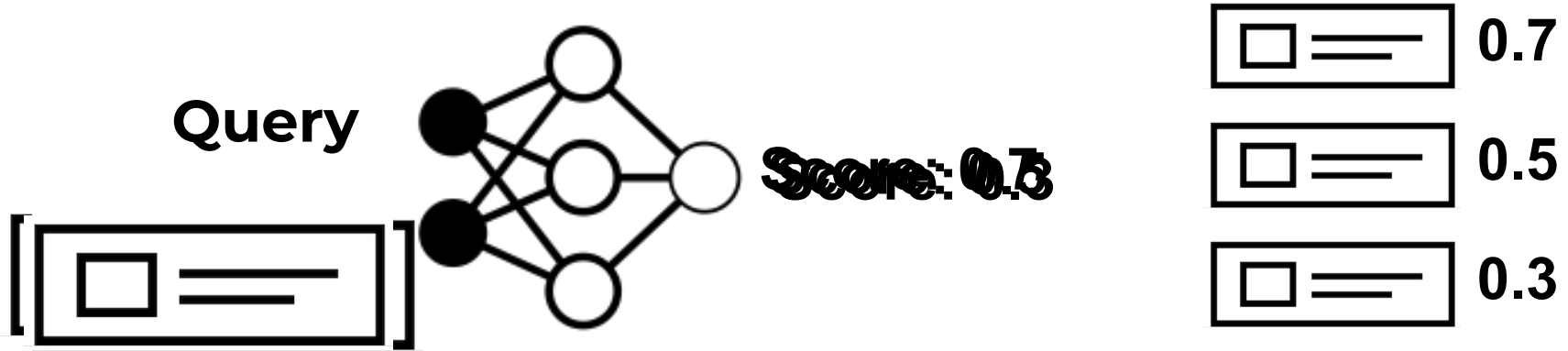
Test E2E

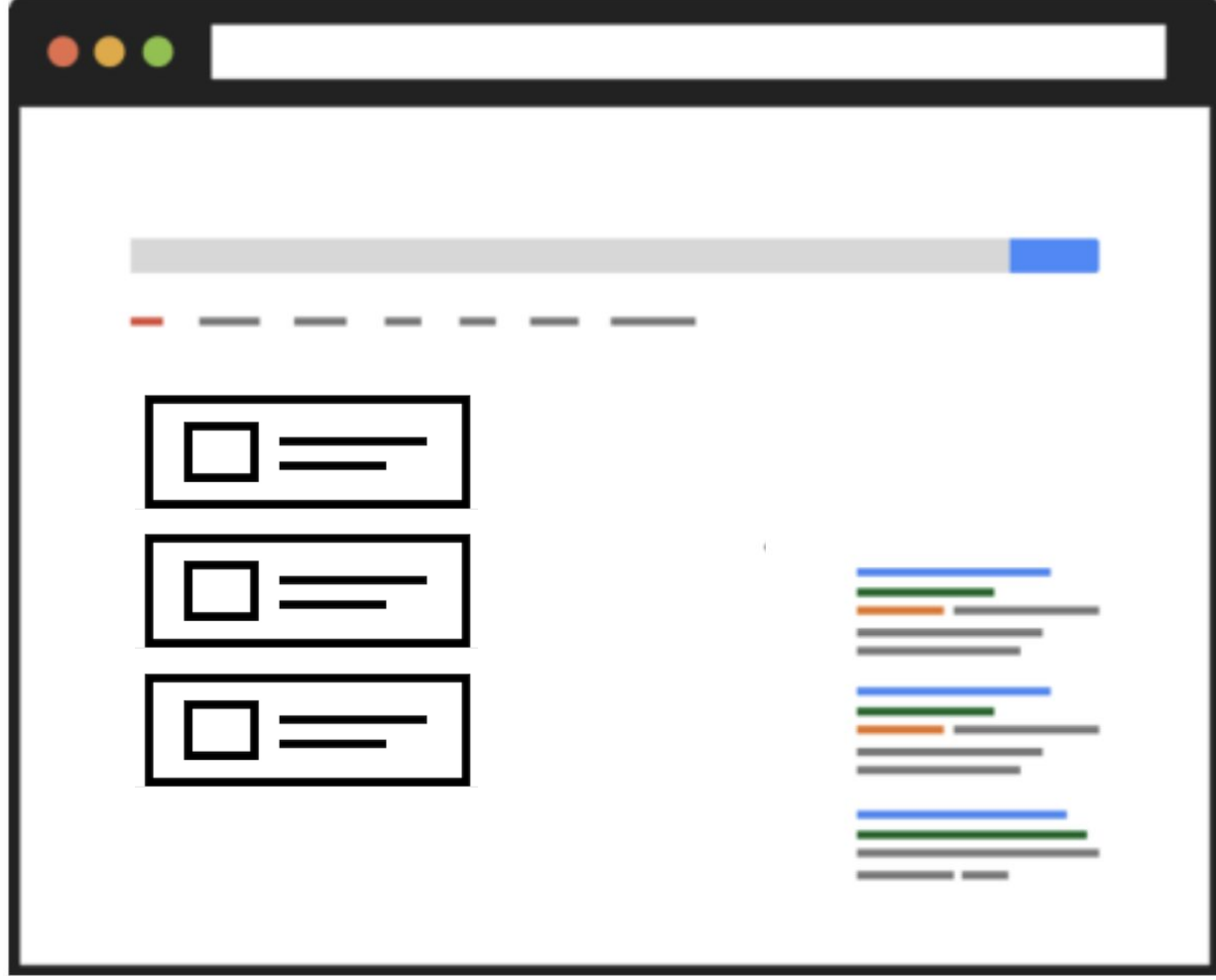
Future

Your Apps



# AI-Based Ranker Overview

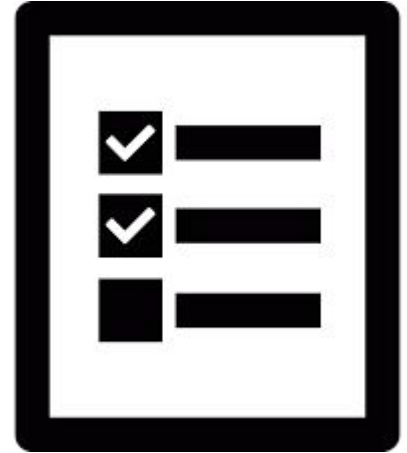
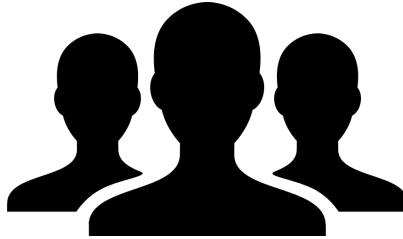
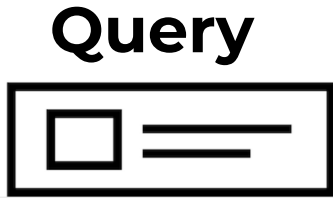






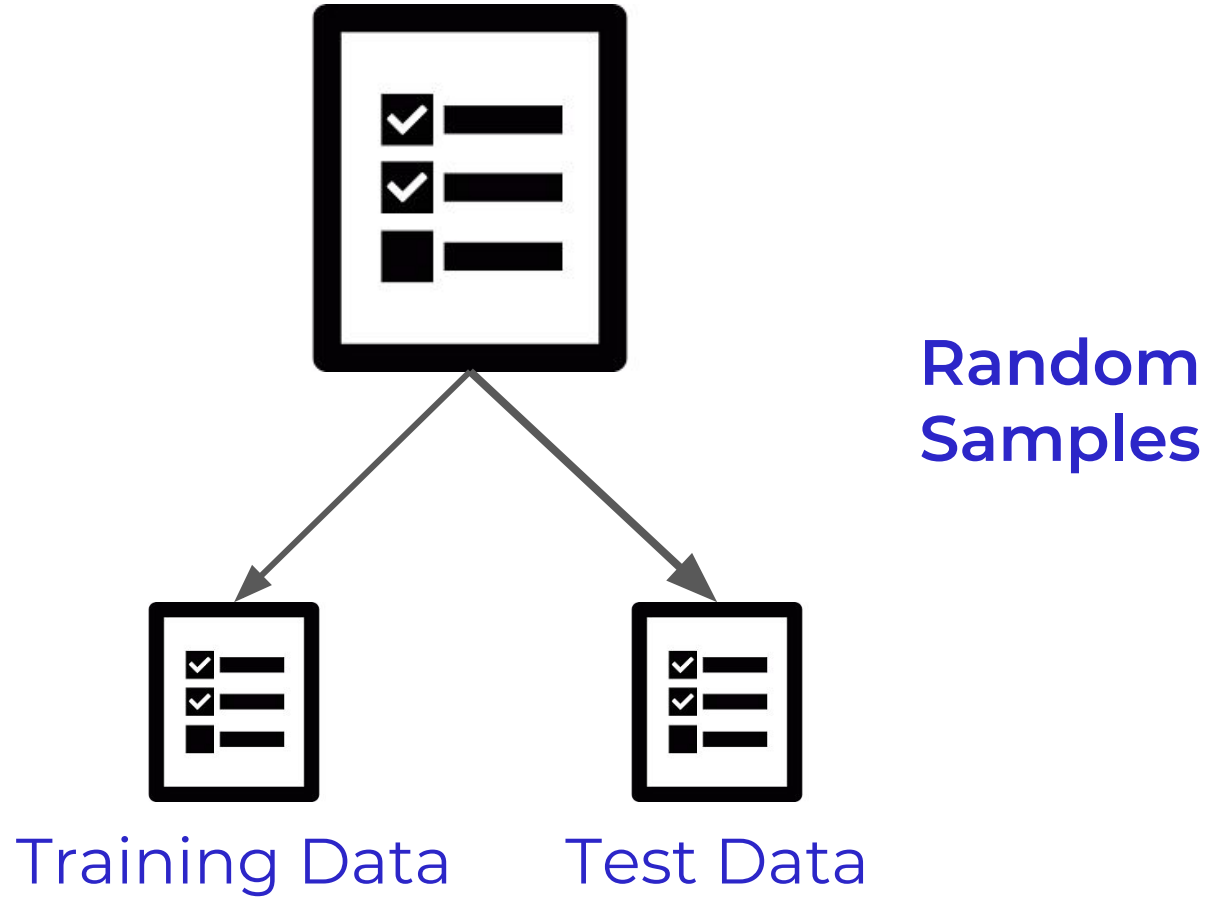


# Training Data



30,000 Queries







## Storytime

“We beat Google in Japan!”





## Storytime

“We tested with the training set...”

“Many top results hard-coded...”

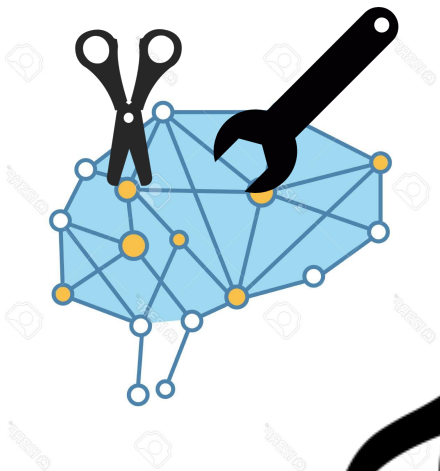


Train

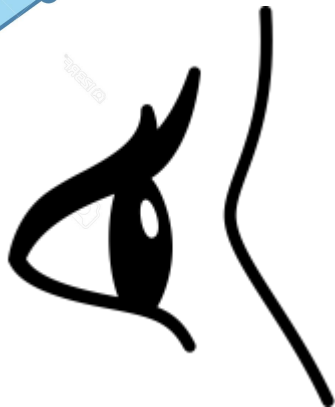
Guess/TrainingSet

Features

Q + Result



**0.9?**



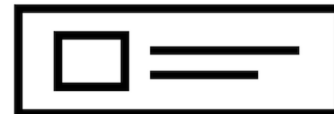
Q Appears # Times



Q in Title?

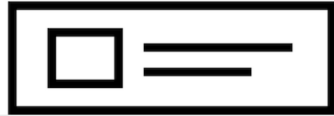


Result Spammy?



**0.3**

# Test Feature Computation



Text



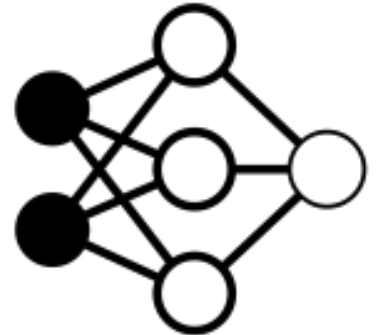
# Query Occurrences?



# Query In Title?



# Result Spammy?



“Microsoft Search Engineers using Google?”



Storytime





# Bias in Training Data

From people getting paid \$20/hour

Not Software Engineers



# Overlap + Confusion

4 People Agree

Q="Bush" Best Result="George Bush"

1 Disagrees

Q="Bush" Best Result="the Band"

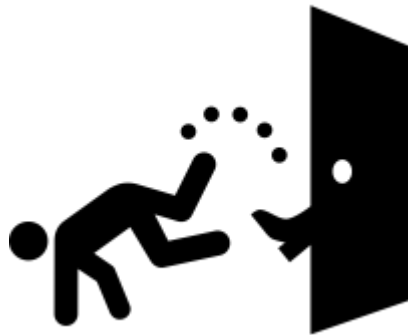


# Overlap + Confusion

Vendor fired the outlier.

No 'right answer'.

Lost Perspective in training Data



# What data don't you have?

Small web index: Can't train on what you don't know





# Training

# Testing AI Training System

Test just like any functional system

Scale

Reporting Accuracy

Look for anomalies...





## Relevance Engineers

Pick/Write Features

Initial Weights, Network Topologies,

Training Variables

- System is Fast
- System is Debuggable/Visible

Celebrated: “One Relevance Engineer beating all others!”



Storytime







## Storytime

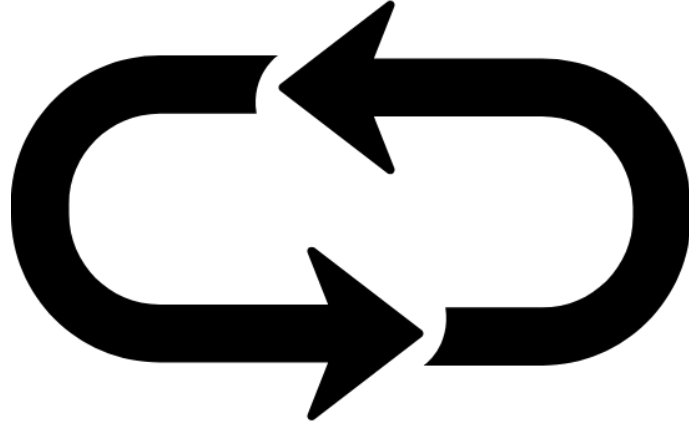
Noticed: One engineer using 80% of compute...





## Storytime

Just kept re-testing system with jitter until he got a winning number.

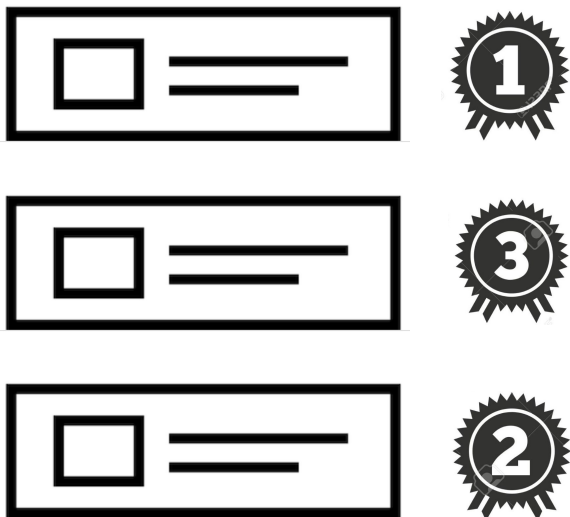




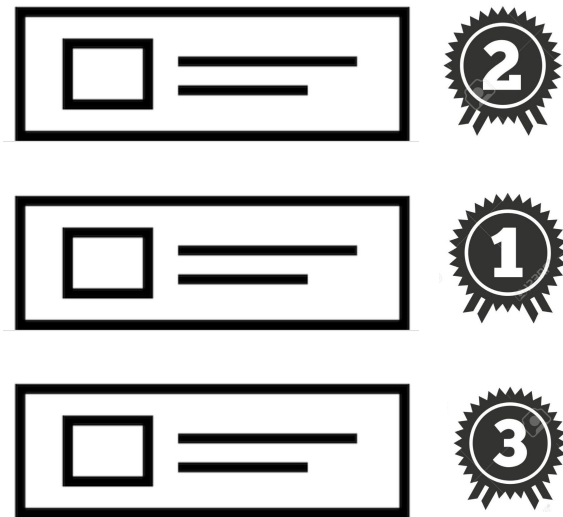
**E2E**

# E2E Result Set Relevance: Discounted Cumulative Gain

Not Good



Far Worse

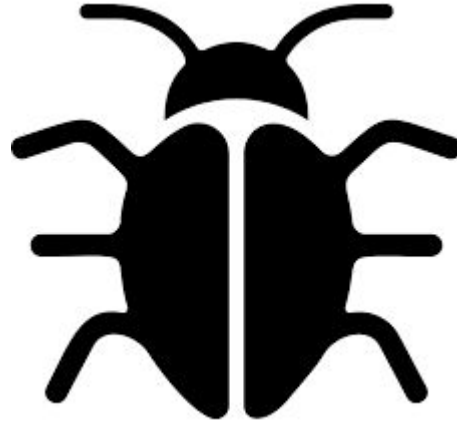




Some % of Web Searches Blank

→ 0 DCG!

Functional Bugs Impact Relevance



# Time To Answer

Ultimate Goal is to Get User an  
Answer, a click.

Seems obvious...





## Storytime

Some Queries are Navigational

- Click one

Some Informational

- User wants to see all top results...

Queries with Many different topics  
take users longer to click a link



## Storytime

Users execute search in tab, but get distracted in the real world.

Switch Tabs

Interrupted

Careful with data...





## Storytime

# Functional Performance Impacts 'Time to Answer'



# Drift

Web Content changes over time

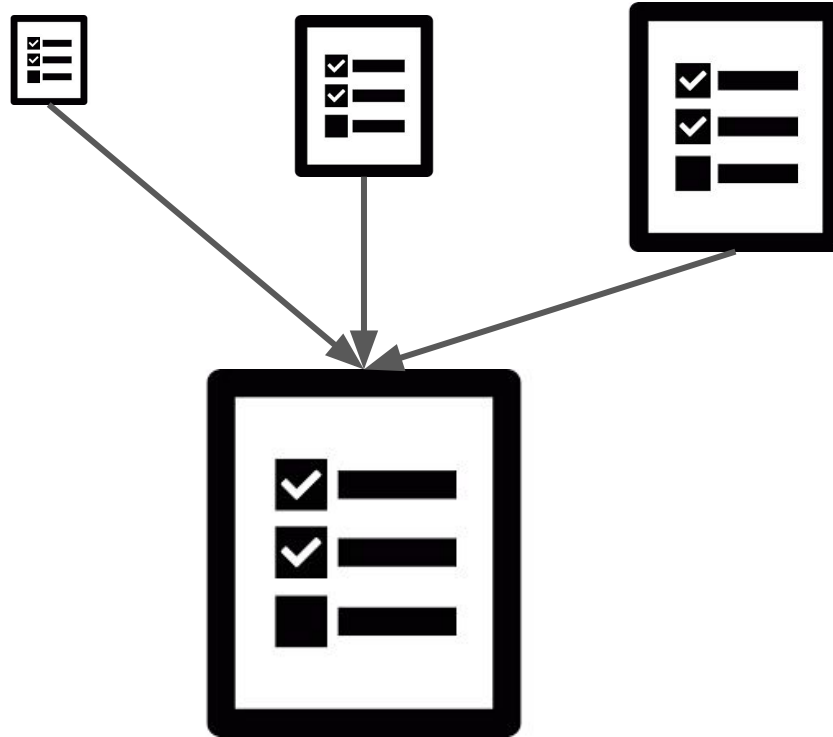
Web Structure changes over time

People's queries change over time



# Drift: Historical Sampling

Last Year Last Month Last Week





## Storytime

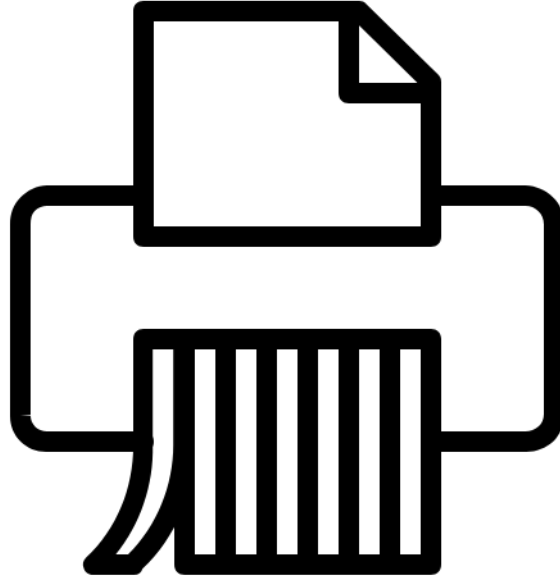
# Internal Feedback on Search Results





## Storytime

Anecdotal Data Almost useless.  
Most ideas help one query, hurt five.

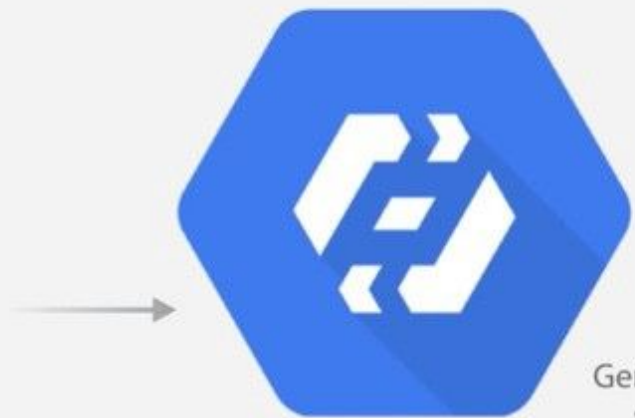




# Future



Photo Dataset



Generate predictions  
with a REST API



Rest API

## Cloud AutoML Vision



Train



Deploy

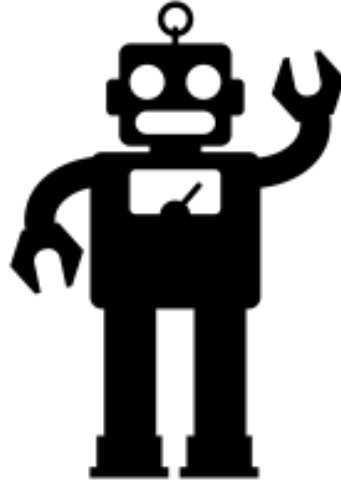


Serve





# Neural Networks Replace Relevance Engineers

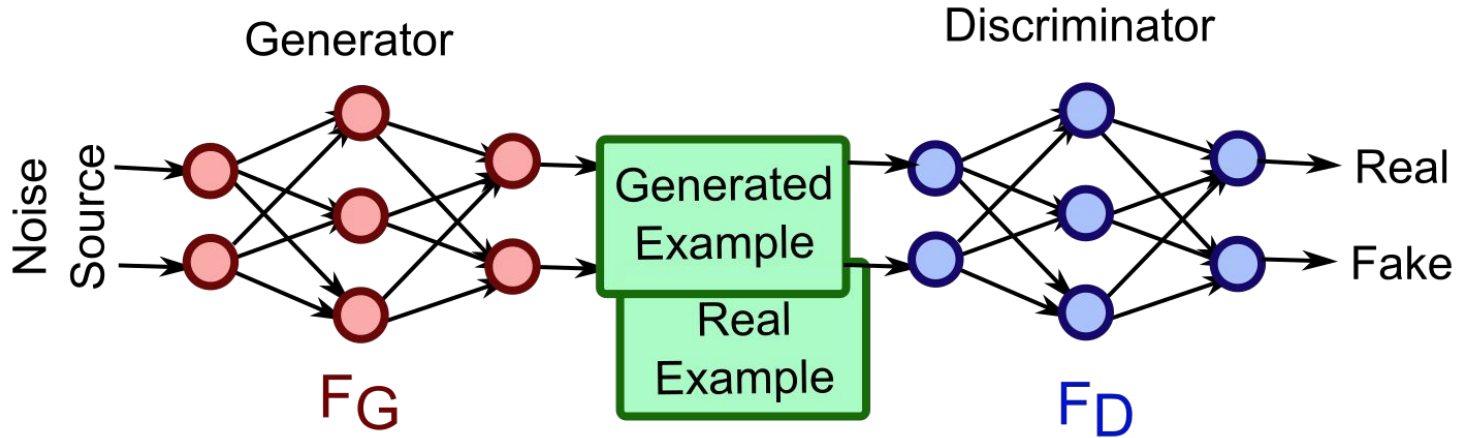






## Storytime

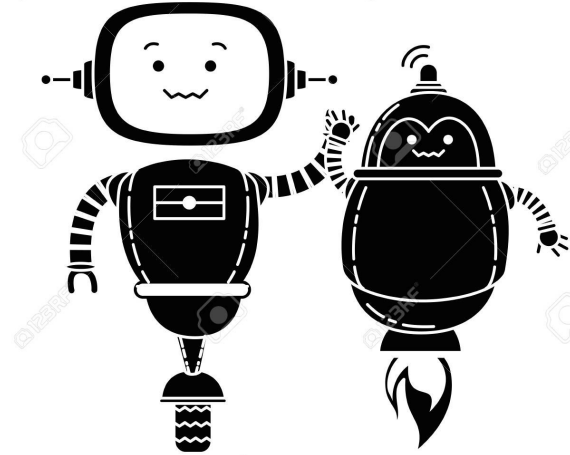
# Generative and Adversarial Networks





## Storytime

Adversarial Networks Don't need humans to test or write the software, just a 'fit() function'.





# AI in Your Apps?

AITesting.org

# AI ST

ARTIFICIAL INTELLIGENCE FOR SOFTWARE TESTING

# WE WANT YOUR FEEDBACK



test.ai

# We're building the future of automated testing

Stay tuned for updates

Get on the List





QnA

jason@test.ai

**Appdiff, Inc.** 340 Brannan St #100, San Francisco, CA 94107, United States

Contents of this presentation are proprietary and confidential. The presentation, or any elements of the presentation may not be duplicated, distributed, shared or used in any way other than for purposes of review, while under NDA. All contents property of Appdiff Inc. unless otherwise assigned to another party. Copyright © Appdiff Inc, 2019.