Nothing to Recommend It

An (Interactive) ML Outage Fable

Todd Underwood @tmu ♦
tmu@google.com ♦
2021-Oct ♦
SRECon ML Track

Agenda

A Tale of Sadness and Woe

An online knitting and crocheting shop is having trouble.

I'll tell the story, you follow along at home trying to guess how to solve it.

There will be a quiz!

- Welcome to Yarnit!
- Weird Recommendations, Less Money
- Same questions, different answers
- What has changed?
- Could this be avoided?
- Lessons

Quick Introduction

ML Outage Fable

Basic Questions (the usual)


Photo by Laårk Boshoff on Unsplash

Who am I?

Recovering Sys/Net engineer, working on ML SRE at Google for 12+ years.

Founded and lead ML SRE and Pittsburgh site for Google.

Writing an Book on Reliable ML for O'Reilly (with a bunch of other people). Some chapters already online. Most importantly: animal decision still pending!

Who are you?

SREs who are curious how ML incidents are different than other kinds of outages.

Why are we here?

To walk through a fake outage and have fun in the process.


The Scene
ML Outage Fable

YarnIt (yarnit.ai)

A Web Shop for All Your Knitting and Crocheting Needs

YarnIt sells yarn. And knitting needles. And patterns.

We use ML in several ways:

- Search results ranking: users search for products and we rank the results using an ML model
- Discovery UX: In various contexts on the site (logged in home page, browsing, cart addition), we use an ML model to present products the customers might be interested in.

A Warm Spring Day

Gabi, ML Engineer, enjoys an unseasonably warm day

Gabi grumbles about climate change, but genuinely enjoys the day.

Two internal escalations (not automated pages) arrive:

- Support says customers are complaining about “weird” search results. The problem reports are vague but have persisted for several days.
- Finance reports that overall revenue for the site is down by a fair bit. They’re not sure how to localize it yet.

The Incident

ML Outage Fable

Audience Participation!

Virtual Audience Participation is Probably Terrible

(I'm trying it anyway)

How it works:

I ask a question. You pause video and write down your answer. Honor system!

The prize? Bragging rights!

Photo by Nicholas Green on Unsplash

Question #1

What Connects The Two Reports?

A. Recently updated (and broken) ML model for recommendations
B. Problem in serving (feature retrieval or training/serving feature definition mismatch)
C. Stale model that hadn't been retrained recently
D. Complex global supply chain problem
E. All of the above, in some combination caused the outage.
F. None of the above. There is no connection.
G. No clue! Not enough information.

ML Outage Fable

The Answer is….

Hold That Thought

Keep your written answer and we will come back to it.

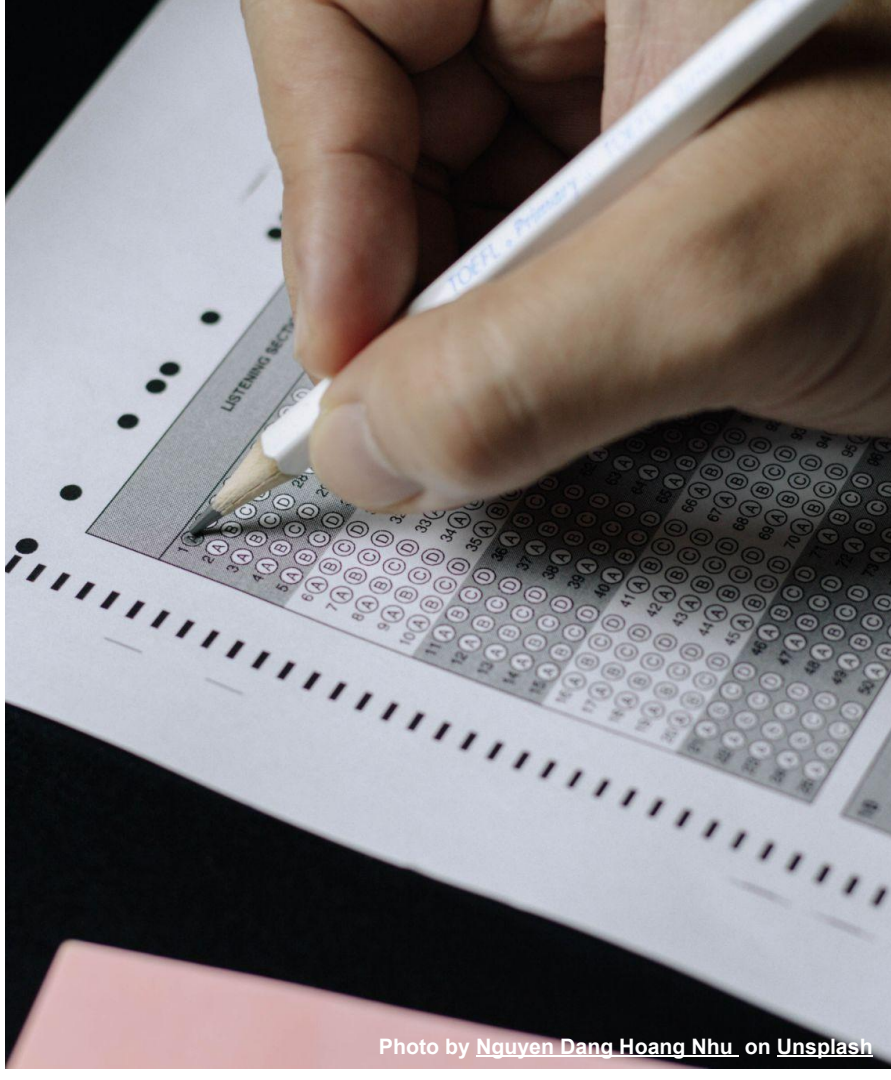


Photo by Nguyen Dang Hoang Nhu on Unsplash

Troubleshooting

Talking to everyone

ML outages often involve more and more diverse parts of an organization than other systems outages.

Gabi talks to:

- Support
- Finance
- Modeling engineers
- SREs
- Retail managers

Early findings:

- Model config is stable, model **has** been trained on new data, no serving errors or training/serving skew.

Troubleshooting

Preliminary:

Probably not:

- A: Broken Model
- B: Serving mishap of some kind
- C: Stale model

Logs

Gabi grabs colleague Imani and they select 100k user queries from today and 100k user queries from four weeks ago.

Question #2

What Will Imani and Gabi Find?

A. The queries are similar but the answers are different. The model has changed!
B. The models are the same but the queries are different.The queries that have changed!
C. The models **and** the queries are different!
D. Everything is the same and this is all a troubleshooting dead end.
E. Grabbing 100k user queries without careful anonymization has privacy implications, a serious problem for the incident investigation.
F. A and E
G. B and E
H. C and E

Photo by Annie Spratt on Unsplash

ML Outage Fable

A Privacy Interlude

ML Outage Fable

ML Incident Response Requires Privacy Infrastructure

E is True: Viewing User Queries is Sensitive

ML model train on data that is often private to users.

ML Production Engineers and Modelers may need to see those data but only in a controlled, logged, and overseen fashion.

Best practices:

- No access to user data by default
- Non emergency access to user data pre-reviewed by privacy experts
- Emergency access to user data logged and post-reviewed by privacy experts
- Anonymization wherever possible.

The Answer...

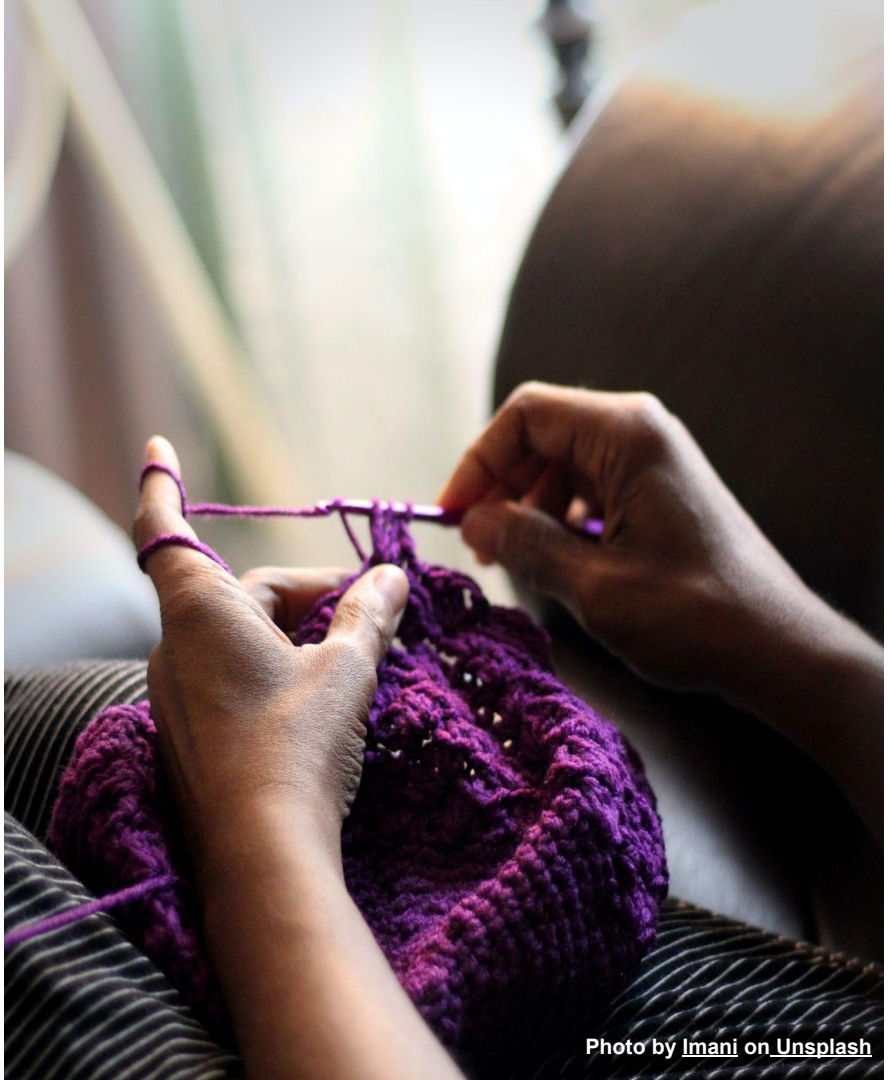
ML Outage Fable

Results!

What "Actually" Happened

1) Temps got warm; user demand shifted to lighter weight yarns.
2) YarnIt's best supplier of lightweight yarns is out of stock of almost everything! (Stupid pandemic supply chain!)
3) Recommendations model makes "weird" recommendations since we cannot recommend product we do not have.
4) Revenue declines since we don't have the products people want to buy!

Not ML-Caused



Correct Answers

Question #1

D: Complex global supply chain problem.
This was not a very plausible guess, but ML problems sometimes have completely exogenous causes.

Question #2

G:B & E: The queries have changed (people want lighter weight yarn) and also we committed some privacy-risky moves by just looking at user queries.

Take Aways

Sometimes ML is Weird
ML systems often have outages typical of other distributed systems, but not always.

Harder to Monitor
Monitoring ML model quality is hard. This is a big topic but it's hard. This means that many incidents are detected first in public

ML Systems Span Whole Organizations
(right now) People deploy ML when it matters, so it tends to span the whole organization.

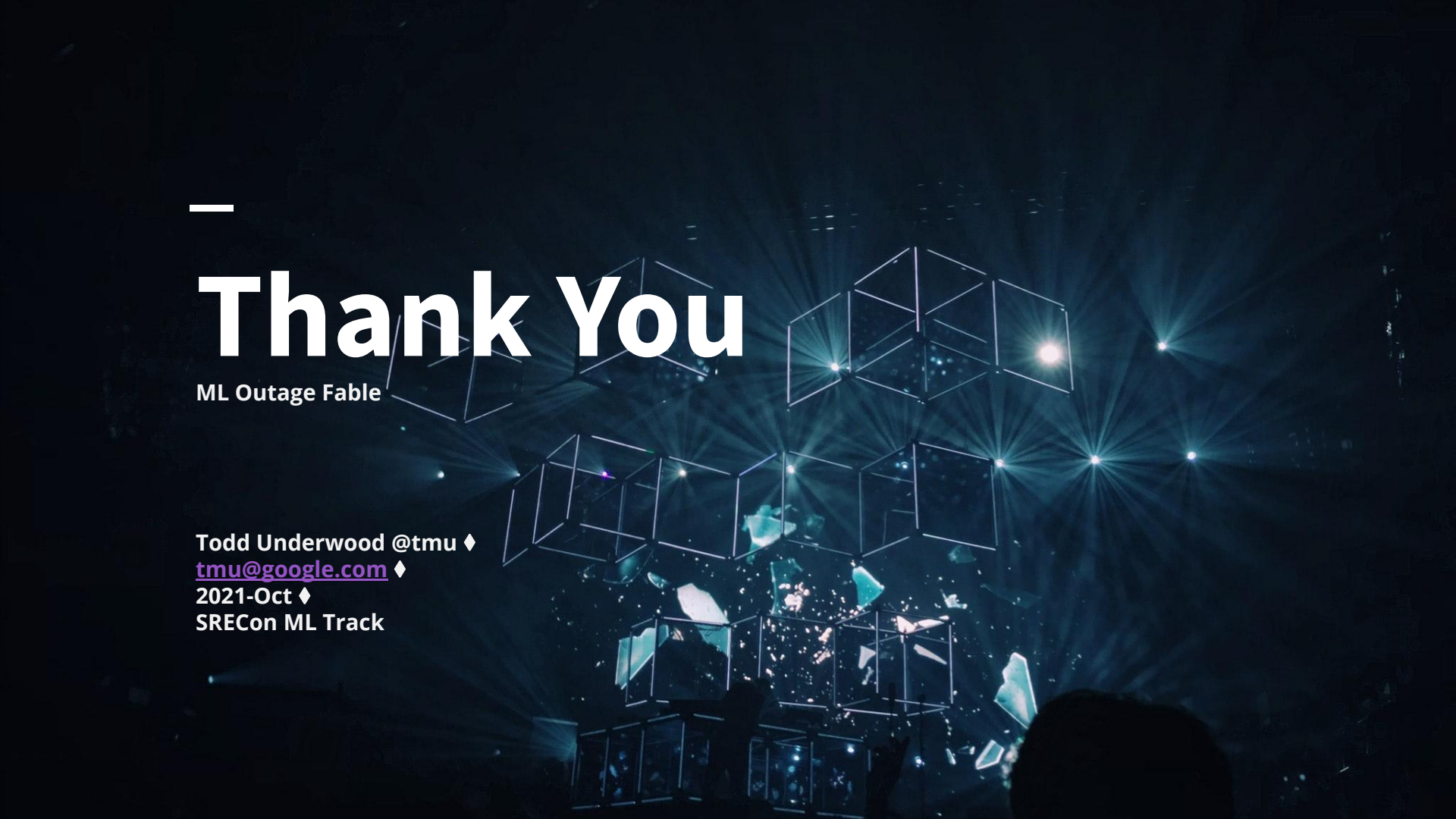
Photo by Maria Lin Kim on Unsplash

Reliable ML book

As mentioned, I'm working on a book with Cathy Chen, Niall Murphy, Kranti Parisa, and the inimitable, one and only D Sculley.

Subscribe to O'Reilly Learning Platform to get early release chapters. Here's a 30-day free code if you want it:
https://learning.oreilly.com/get-learning/?code=RML22

Code is good for 30 days once you subscribe. The link expires Dec/31/22.



Thank You

ML Outage Fable

Todd Underwood @tmu ♦
tmu@google.com ♦
2021-Oct ♦
SRECon ML Track