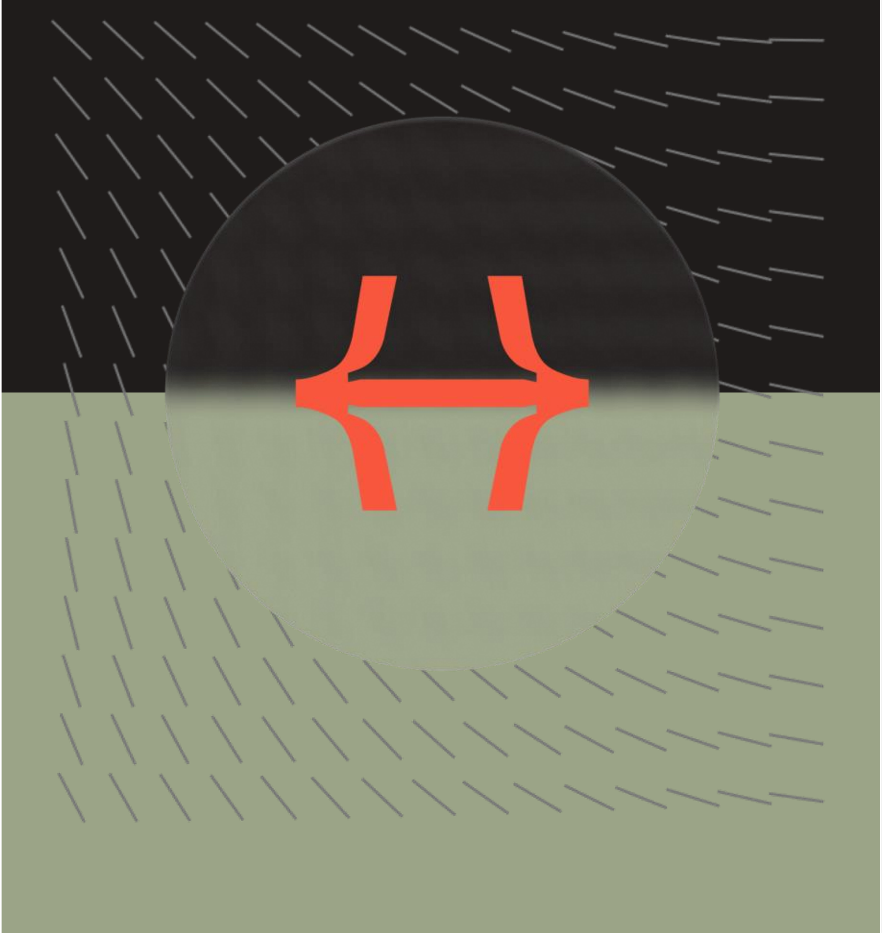
Presentation: Testing Assembly Seminar 11/2024

Business Continuity in an Unpredictable World

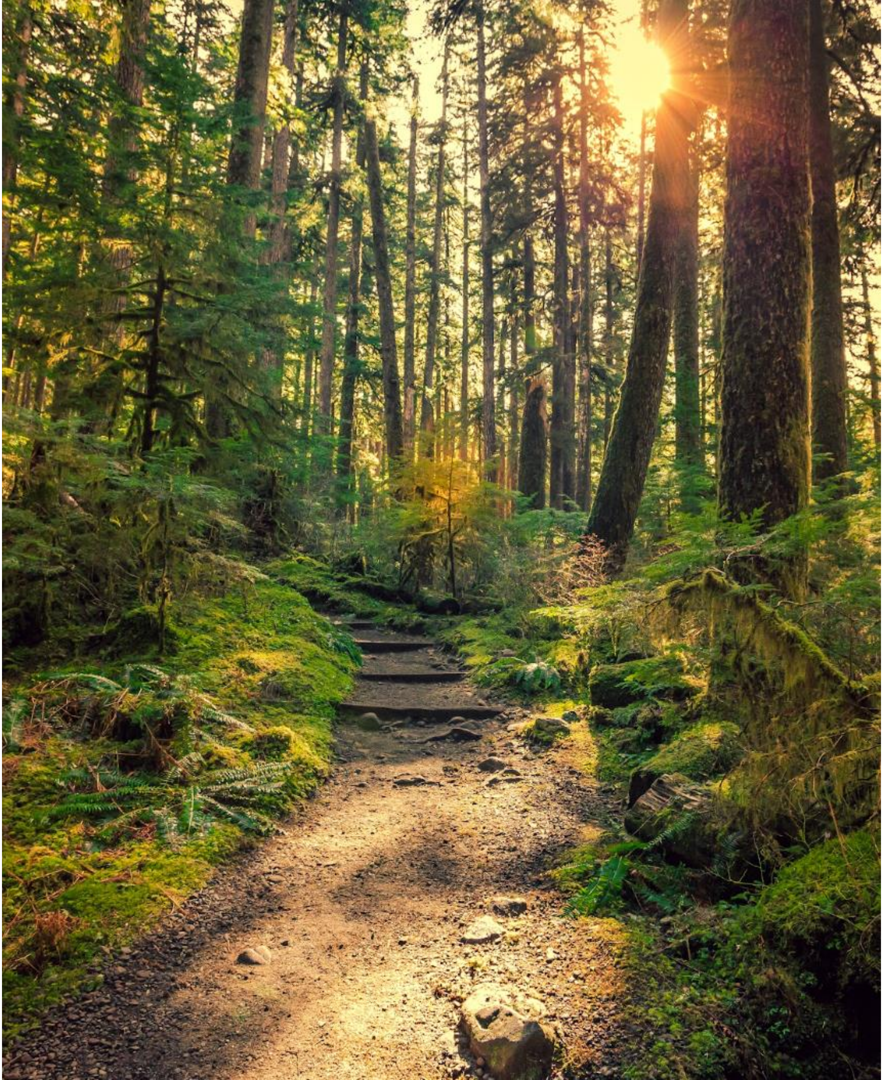
HIDDEN TRAIL



Hi there, I’m Tuomas!

- A lifetime in the digital business
- QA entrepreneur for a decade+
- DevOps side quest for 5+ years
- Guiding thought:

  Focusing on quality and experience (both users and creators) spawns productivity



What to expect

We will walk through definitions, scenarios and example stories.

Focus in this presentation is on ensuring business continuity in digital services landscape.

~35 minutes for presentation, then some for discussion.

Contents

Setting the scene

What do we mean with unpredictability?
How to define business continuity?

Operations / Production resiliency

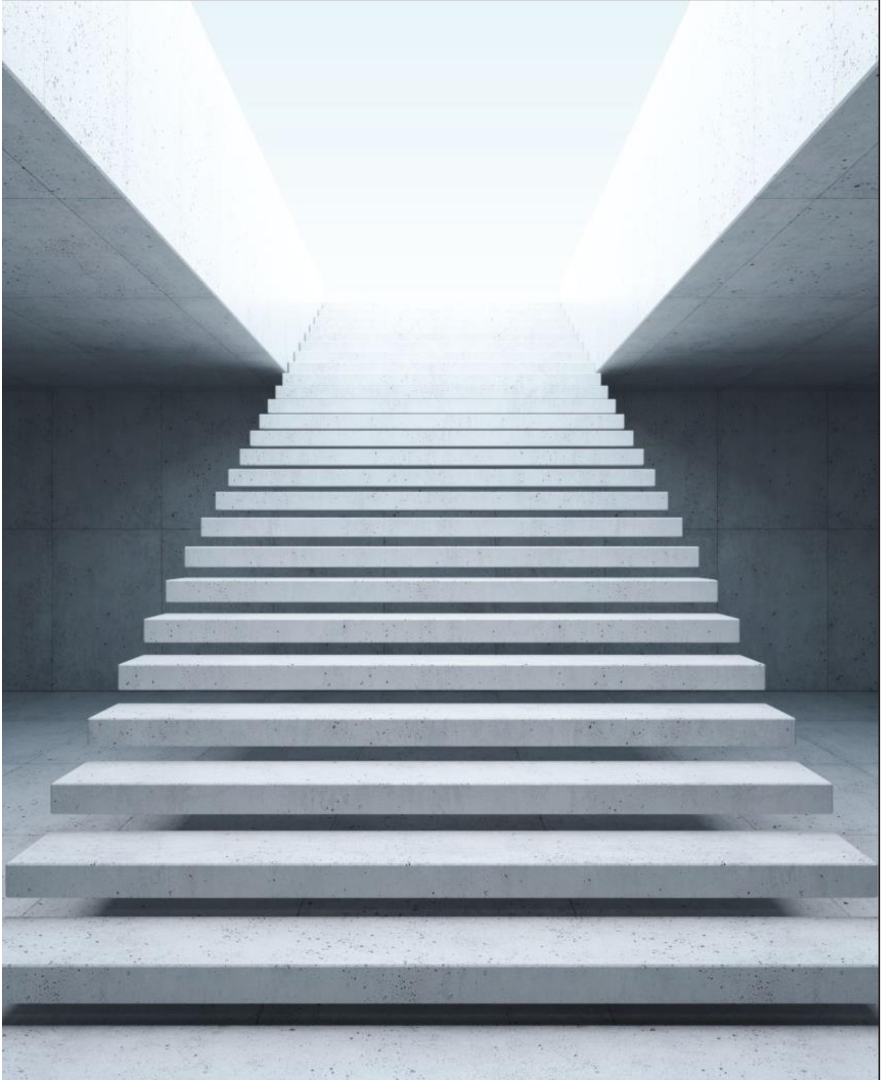
Quality in software operations -
How to improve resiliency in practice?

Case Study: Generative AI

Production-grade delivery of
AI-powered digital services.

Quality in the Unpredictable World

Discuss quality assurance and engineering as
professional capabilities



Setting the scene

“At its best, life is completely unpredictable.”

-Christopher Walken

Business Continuity?

Technology aspect:

- Ensure operational readiness
- Avoid downtime
- Quickly restore from failure

Business & innovation aspect:

- Inspire loyalty
- Keep ahead of the curve
- Know your landscape
- Grow with your creators

Unpredictable world?

Technology aspect:

- “Intelligent”, increasingly complex systems
- Hybrid connectivity, environments and diversity of devices
- Data volume, privacy and evolving legal landscape

Business & innovation aspect:

- Demand for personalization and tailored experiences
- Shift to new business models (e.g. subscriptions, data-driven)
- Expansion and loyalty regarding digital ecosystems

Erosion of Trust

**General issue in society
-
‘Post-truth era’.**

Confusion and machine-in-the-loop can lead to losing the foundation for collaboration.

Overpromising and frequent delay feed into the erosion.

Calibrate collisions between product and tech.

Transparency is a two-edged sword.

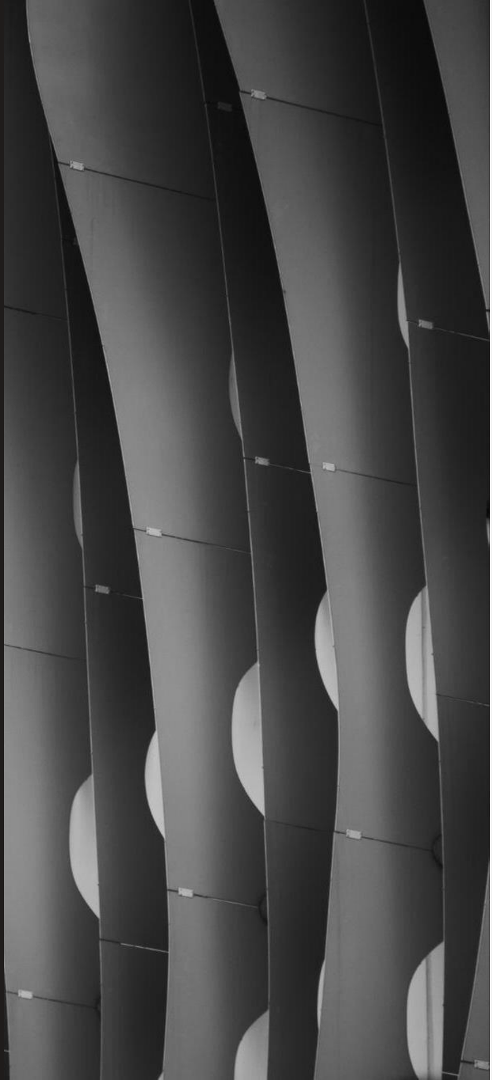
Opacity can shield cognitive focus.

Lack of general accountability.

Taking active responsibility cascades.

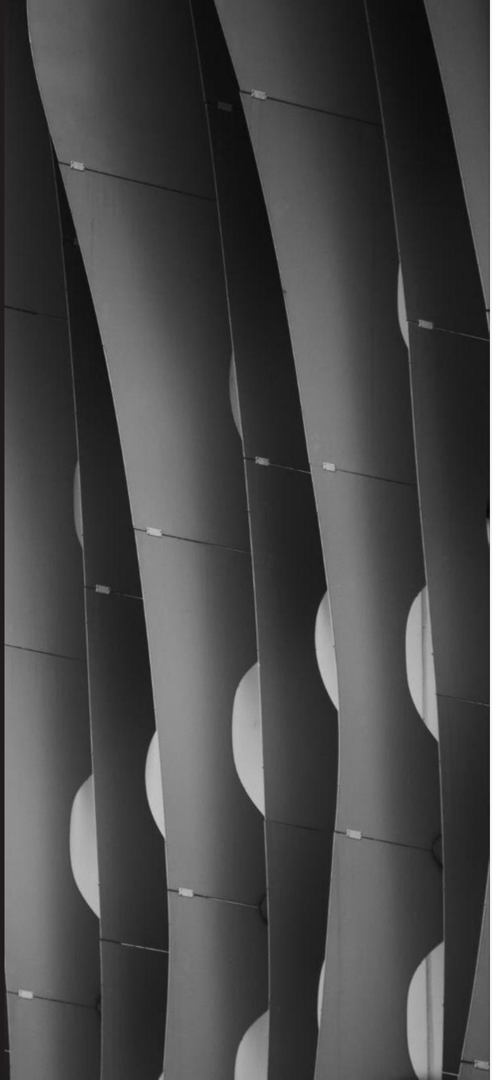
Common dangers

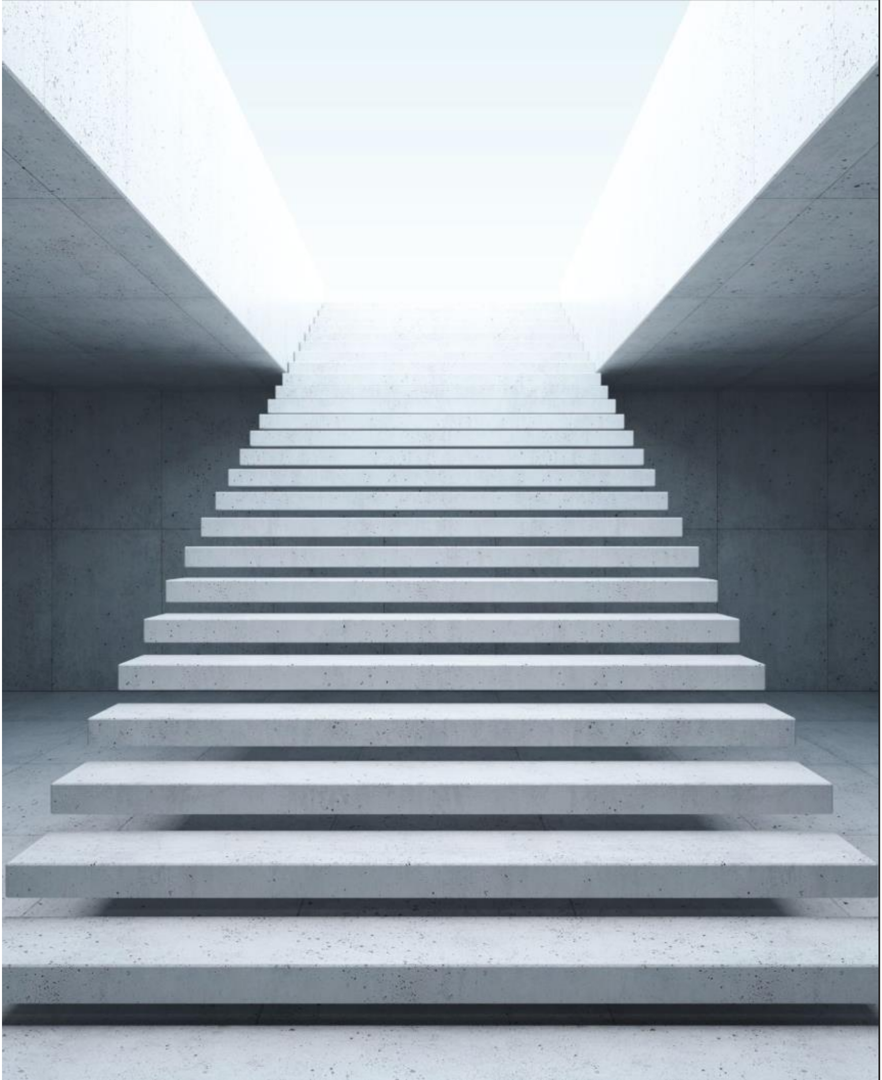
- **Preparedness Paradox:** Cognitive biases and short-term work prioritization
- **Inadequate Testing:** Disaster recovery and business continuity exercises
- **Gaps in Planning:** Cover all essential IT systems and disaster scenarios
- **Unclear Roles and Responsibilities:** Poorly defined roles and insufficient training



Impact & price

- **Downtime Costs**: Revenue loss due to halted operations, missed transactions, and lost business opportunities
- **Recovery Expenses**: Unplanned restore, repair and emergency IT services
- **Regulatory Fines**: Non-compliance with industry regulations (e.g., data protection laws)
- **Contractual Penalties**: SLAs or other contractual obligation breaches lead to penalties or compensation payments



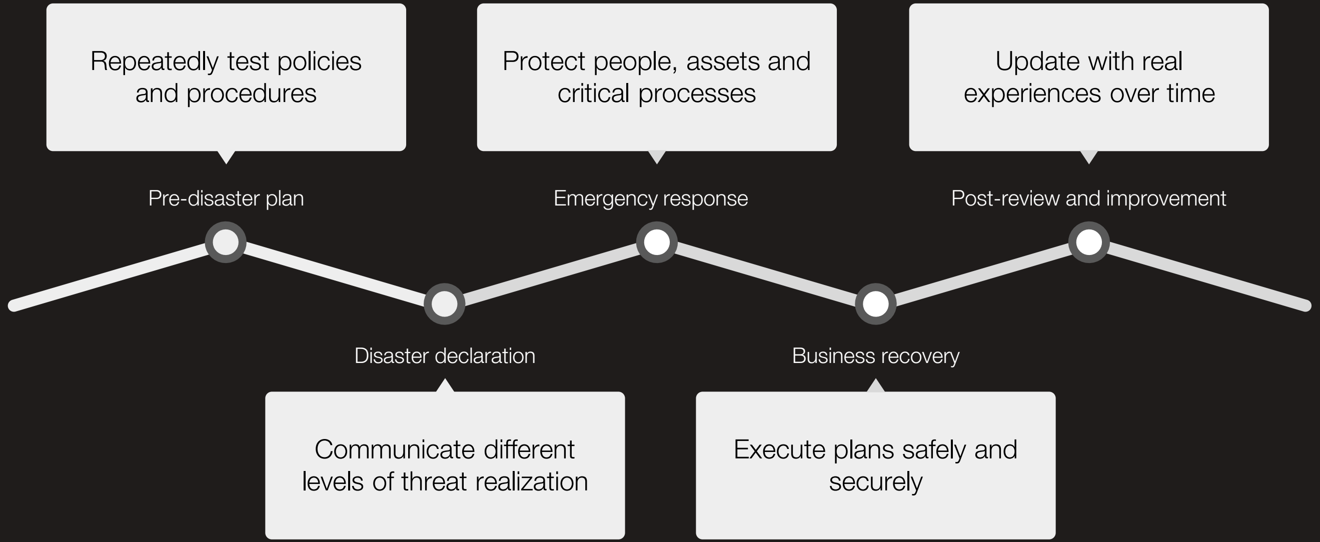


Operations / Production Resiliency

Business Continuity Key Artifacts

- Risk assessment culture
  - Potential threats and likelihood
- Business Impact Analysis (BIA)
  - Severity and impact to finances etc.
- Business Continuity Plan(s)
  - Strategic frameworks for recovering from disruptions
- Crisis management readiness
  - Communications playbook
- Disaster Recovery Planning (DRP)
  - Practical operations playbook

Business Continuity Plan



Repeatedly test policies and procedures

Pre-disaster plan

Disaster declaration

Communicate different levels of threat realization

Protect people, assets and critical processes

Emergency response

Business recovery

Execute plans safely and securely

Update with real experiences over time

Post-review and improvement

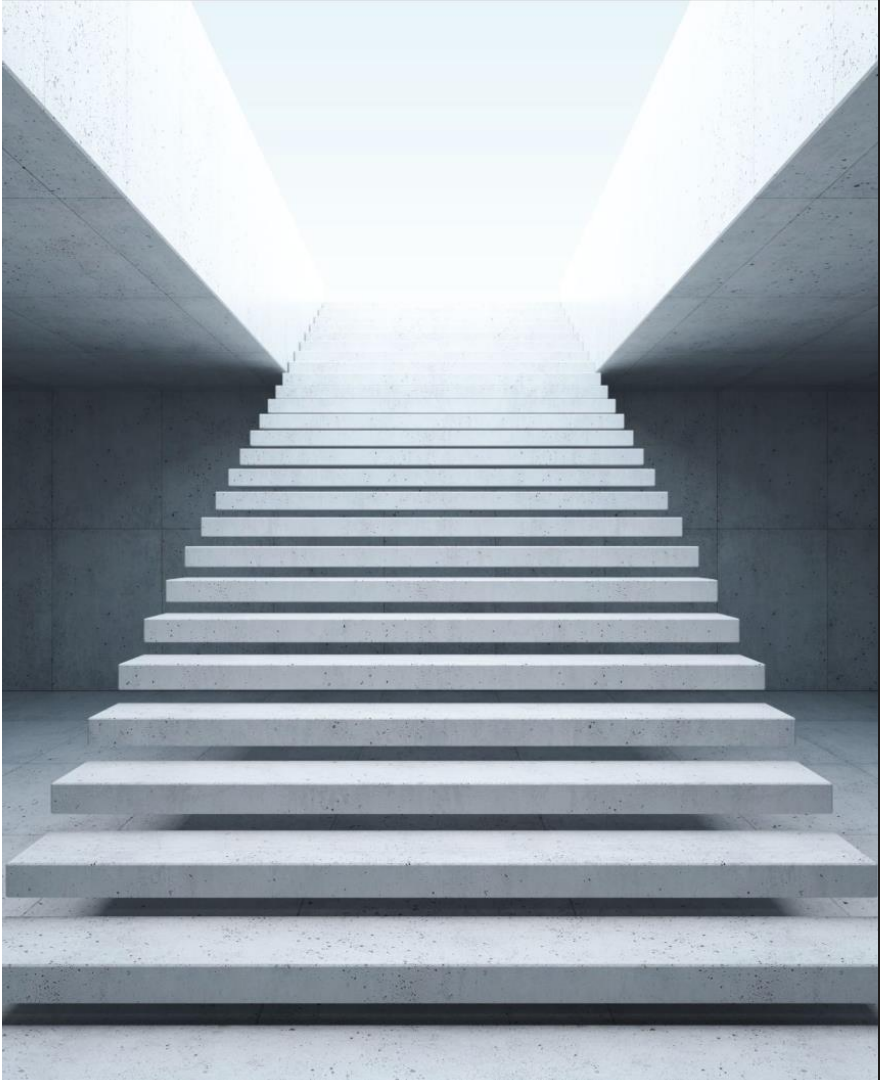
Operational resiliency and capability testing

- Log monitoring and validation across systems
- Alerts, including conditions and routing
- Observability, metrics and distributed tracing of application data
- Incident response, communication and responsibilities
- Failover, redundancy, restoration and disaster recovery
- Self-healing, capacity and automated processes

Monitoring & Alert fatigue

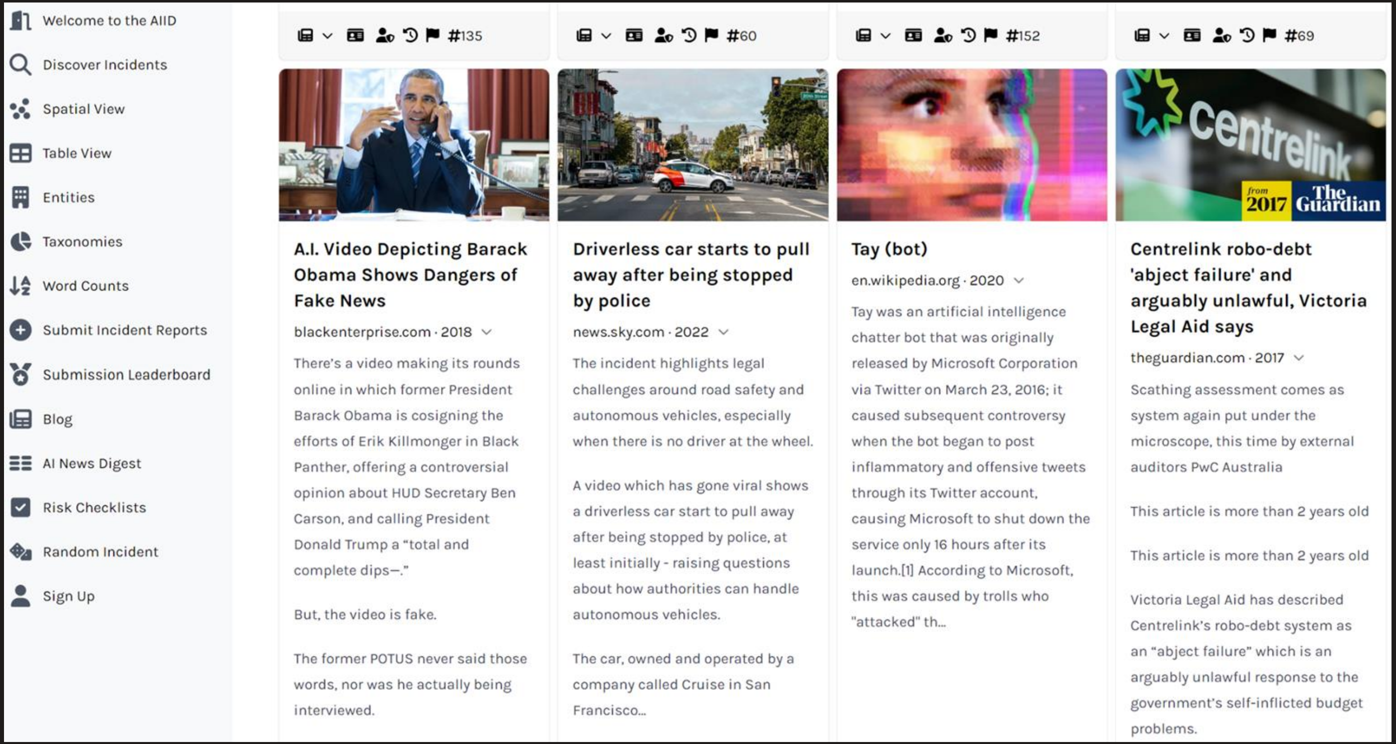
- Majority of alerts tend to be ignored
- Alerts are normalized through repetition
- Avoid sharing alerts to teams outside of responsibility area ‘just in case’

- Each alert should be actionable
- Active review of alert levels and thresholds
- Use human cues - sensory and visual



Case Study: Production-grade delivery of AI-powered services

Incidentdatabase.ai



A.I. Video Depicting Barack Obama Shows Dangers of Fake News

blackenterprise.com · 2018

There's a video making its rounds online in which former President Barack Obama is cosigning the efforts of Erik Killmonger in Black Panther, offering a controversial opinion about HUD Secretary Ben Carson, and calling President Donald Trump a "total and complete dips—."

But, the video is fake.

The former POTUS never said those words, nor was he actually being interviewed.

Driverless car starts to pull away after being stopped by police

news.sky.com · 2022

The incident highlights legal challenges around road safety and autonomous vehicles, especially when there is no driver at the wheel.

A video which has gone viral shows a driverless car start to pull away after being stopped by police, at least initially - raising questions about how authorities can handle autonomous vehicles.

The car, owned and operated by a company called Cruise in San Francisco...

Tay (bot)

en.wikipedia.org · 2020

Tay was an artificial intelligence chatter bot that was originally released by Microsoft Corporation via Twitter on March 23, 2016; it caused subsequent controversy when the bot began to post inflammatory and offensive tweets through its Twitter account, causing Microsoft to shut down the service only 16 hours after its launch.[1] According to Microsoft, this was caused by trolls who "attacked" th...

Centrelink robo-debt 'abject failure' and arguably unlawful, Victoria Legal Aid says

theguardian.com · 2017

Scathing assessment comes as system again put under the microscope, this time by external auditors PwC Australia

This article is more than 2 years old

This article is more than 2 years old

Victoria Legal Aid has described Centrelink's robo-debt system as an "abject failure" which is an arguably unlawful response to the government's self-inflicted budget problems.

Case 1: “Dear Doc” Application

Usage scenario:

Young people look for personalized, empathetic and actionable help for social and emotional challenges online.

Quality challenges:

- Empathy, tone and toxicity
- Context sensitivity
- Trust and safety
- Engagement and retention

Solutions:

- Use mix of evaluation metrics and NLP
- Memory & dynamic model switching
- Red teaming the application
- Gamify and test for freshness

Case 2: Summarization assistant

Usage scenario:

Tech researchers use an AI-powered assistant to summarize papers from arXiv and other sources.

Quality challenges:

- Factual accuracy
- Domain adaptation
- Bias in summarization
- Scalability

Solutions:

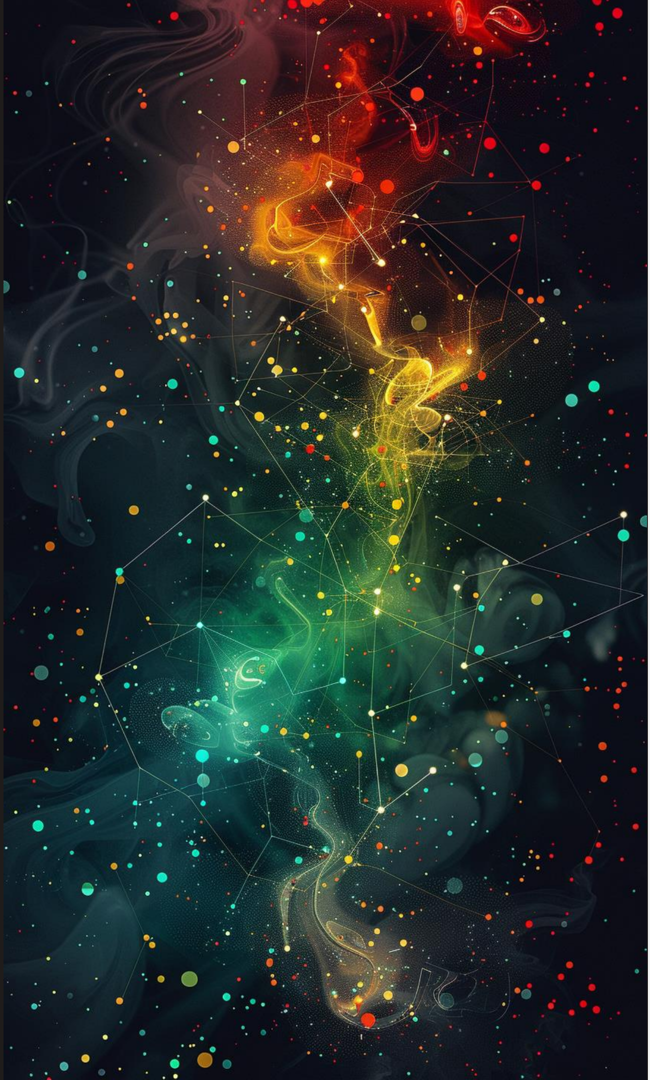
- RAG based references
- Test for terminology and jargon
- Evaluation vs curated dataset
- Divide-and-conquer

Aspect	“Dear Doc” Chatbot	Summarization Assistant
Primary User Need	Empathetic support and guidance	Accurate, concise academic summaries
Critical Quality Focus	Presence, trustworthiness, safety	Factual accuracy, domain specificity
Key Risk	Harmful or disengaging responses	Misrepresentation of academic content
Scalability Challenge	Context management for many active, growing users	Processing large volumes of complex text

Gen AI / LLM Testing

Security & Privacy

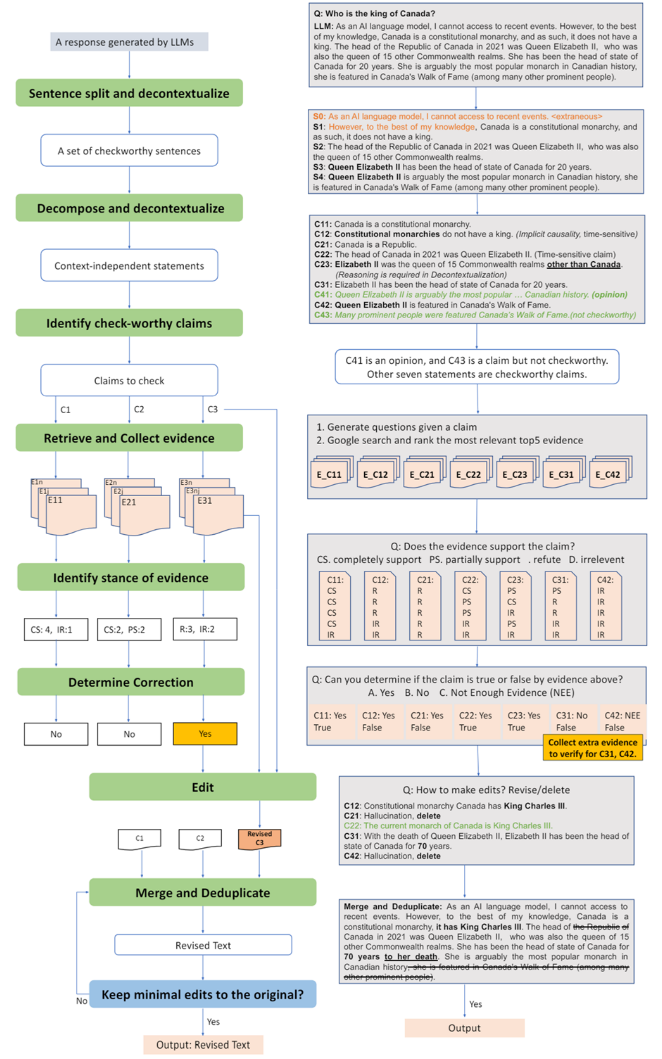
- Functional testing
  - Prompts, behavior and feature coverage
- Performance & cost testing
  - Model comparison (e.g. latest GPT vs local Llama)
  - Token usage optimization
- Quality testing
  - Reasoning, logic
  - Factuality, hallucinations, consistency
  - Bias, fairness, ethicality
- Adversarial testing
  - Robustness, safety
  - Prompt injection, manipulative use
- System testing
  - Agent architectures
  - User experience
  - Total cost & performance
  - Maintainability



New testing paradigms

- AI tends to make information up
- Tools aim to satisfy the users
- LLM creators do not actively take responsibility
- Safeguards must include reasoning steps
- Probabilistic testing needs ranges

https://arxiv.org/pdf/2311.09000


A response generated by LLMs
Sentence split and decontextualize
A set of checkworthy sentences
Decompose and decontextualize
Context-independent statements
Identify check-worthy claims
Claims to check
C1
C2
C3
Retrieve and Collect evidence
E1n
E1j
E11
E2n
E2j
E21
E3n
E3nj
E31
Identify stance of evidence
CS: 4, IR:1
CS:2, PS:2
R:3, IR:2
Determine Correction
No
No
Yes
Edit
C1
C2
Revised C3
Merge and Deduplicate
Revised Text
Keep minimal edits to the original?
No
Yes
Output: Revised Text

Q: Who is the king of Canada?

LLM: As an AI language model, I cannot access to recent events. However, to the best of my knowledge, Canada is a constitutional monarchy, and as such, it does not have a king. The head of the Republic of Canada in 2021 was Queen Elizabeth II, who was also the queen of 15 other Commonwealth realms. She has been the head of state of Canada for 20 years. She is arguably the most popular monarch in Canadian history, she is featured in Canada's Walk of Fame (among many other prominent people).

S0: As an AI language model, I cannot access to recent events. <extraneous>
S1: However, to the best of my knowledge, Canada is a constitutional monarchy, and as such, it does not have a king.
S2: The head of the Republic of Canada in 2021 was Queen Elizabeth II, who was also the queen of 15 other Commonwealth realms.
S3: **Queen Elizabeth II** has been the head of state of Canada for 20 years.
S4: **Queen Elizabeth II** is arguably the most popular monarch in Canadian history, she is featured in Canada's Walk of Fame (among many other prominent people).

C11: Canada is a constitutional monarchy.
C12: **Constitutional monarchies** do not have a king. *(Implicit causality,* time-sensitive*)*
C21: Canada is a Republic.
C22: The head of Canada in 2021 was Queen Elizabeth II. (Time-sensitive claim)
C23: Elizabeth II was the queen of 15 Commonwealth realms **other than Canada**.
(Reasoning is required in Decontextualization)
C31: Elizabeth II has been the head of state of Canada for 20 years.
C41: *Queen Elizabeth II is arguably the most popular … Canadian history.* ***(opinion)***
C42: Queen Elizabeth II is featured in Canada's Walk of Fame.
C43: *Many prominent people were featured Canada's Walk of Fame.(not checkworthy)*

C41 is an opinion, and C43 is a claim but not checkworthy.
Other seven statements are checkworthy claims.

1. Generate questions given a claim
2. Google search and rank the most relevant top5 evidence

E_C11 **E_C12** **E_C21** **E_C22** **E_C23** **E_C31** **E_C42**

Q: Does the evidence support the claim?
CS. completely support PS. partially support . refute D. irrelevent

C11:	C12:	C21:	C22:	C23:	C31:	C42:
CS	R	R	CS	PS	PS	IR
CS	R	R	PS	CS	R	IR
CS	R	R	PS	IR	R	IR
CS	IR	R	IR	PS	IR	IR
IR	IR	R	IR	IR	IR	IR

Q: Can you determine if the claim is true or false by evidence above?
A. Yes B. No C. Not Enough Evidence (NEE)

C11: Yes	C12: Yes	C21: Yes	C22: Yes	C23: Yes	C31: No	C42: NEE
True	False	False	True	True	False	False

Collect extra evidence to verify for C31, C42.

Q: How to make edits? Revise/delete

C12: Constitutional monarchy Canada has **King Charles III**.
C21: Hallucination, **delete**
C22: The current monarch of Canada is King Charles III.
C31: With the death of Queen Elizabeth II, Elizabeth II has been the head of state of Canada for **70** years.
C42: Hallucination, **delete**

Merge and Deduplicate: As an AI language model, I cannot access to recent events. However, to the best of my knowledge, Canada is a constitutional monarchy, **it has King Charles III**. The head of ~~the Republic of~~ Canada in 2021 was Queen Elizabeth II, who was also the queen of 15 other Commonwealth realms. She has been the head of state of Canada for **70 years** **to her death**. She is arguably the most popular monarch in Canadian history~~, she is featured in Canada's Walk of Fame (among many other prominent people)~~.

Yes

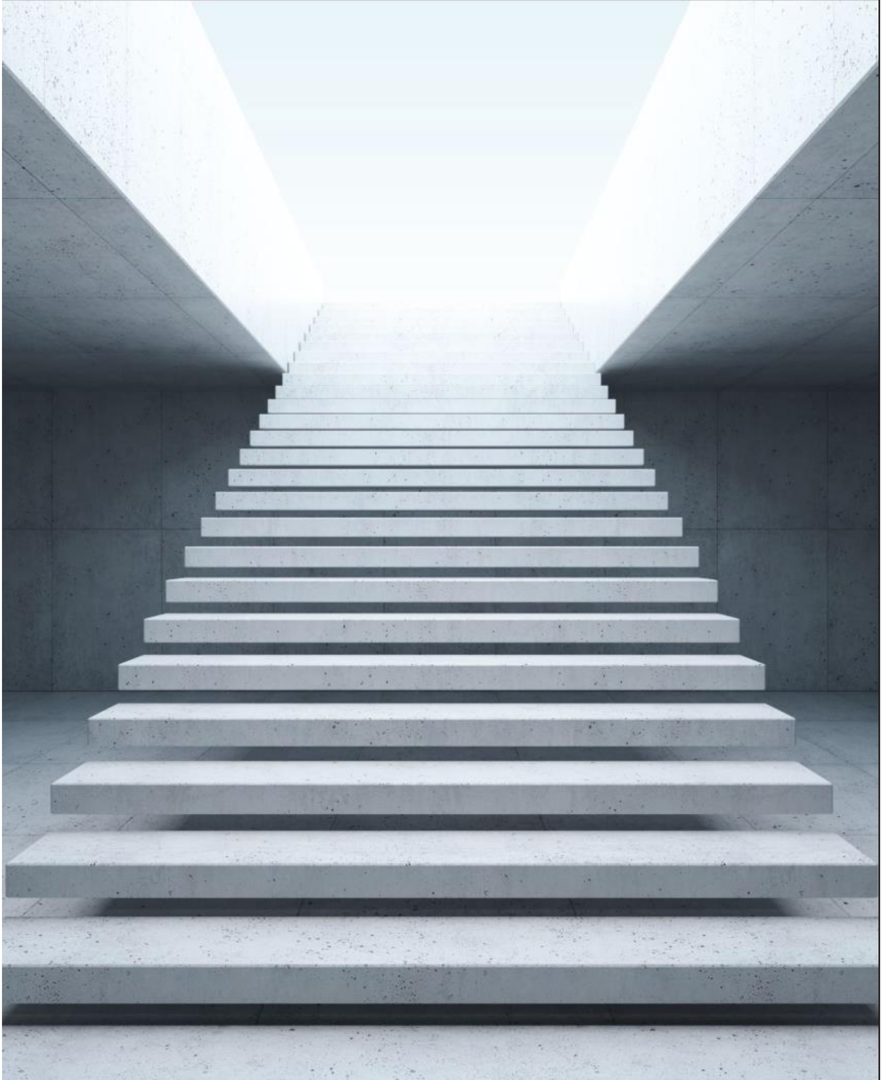
Output

Red Teaming Gen AI Powered Solutions

- Complexity of LLM models are proportional to their output space
- Conjure malicious attacks and harmful behavior

    - *Illegal activities*
    - *Brand image*
    - *Data privacy*
    - *Unauthorized access*
    - *Responsible AI*

- Seek to trigger vulnerabilities by re-framing prompts
- Comprehensive dataset needed
- Scale in all directions
- Specialized strategies



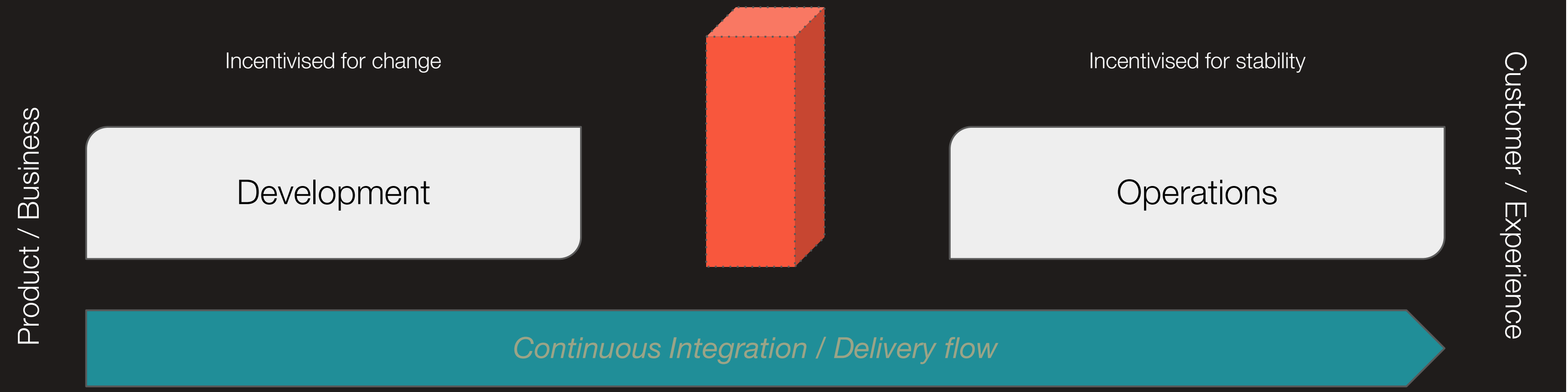
QA in The Unpredictable World

Quality Assurance - Big picture

- QA lays the foundation to risk aversion and management
- QA activities are very efficient in (pre-)development phase
- Operations area quality is rarely considered part of QA focus
- Expectations toward development teams and QA grow all the time

Questions:

- What to do in the face of the changing industry?
- How to take ownership of the narrative regarding QA professionals?



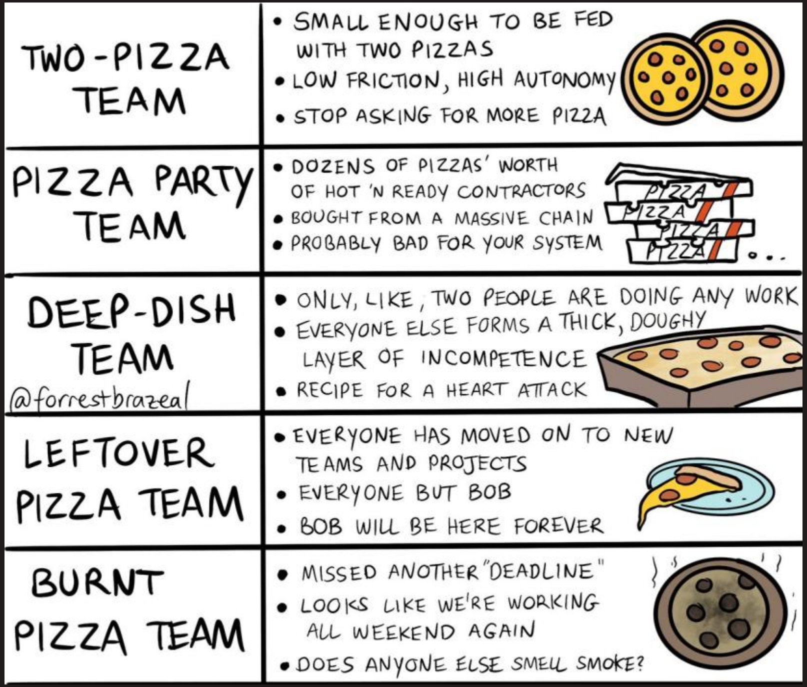
QA has spread to all areas of the software lifecycle. We need to deal with different kinds of drivers.

In a utopian world, everyone would have the same quality goals.

Machine generated work impact on skills and capabilities

What will happen to the professions and skill levels in this industry when various mechanisms allow for reaching satisfactory outputs faster and with less time for introspection?

Team structures



TWO-PIZZA TEAM	• SMALL ENOUGH TO BE FED WITH TWO PIZZAS • LOW FRICTION, HIGH AUTONOMY • STOP ASKING FOR MORE PIZZA
PIZZA PARTY TEAM	• DOZENS OF PIZZAS' WORTH OF HOT 'N READY CONTRACTORS • BOUGHT FROM A MASSIVE CHAIN • PROBABLY BAD FOR YOUR SYSTEM
DEEP-DISH TEAM	• ONLY, LIKE, TWO PEOPLE ARE DOING ANY WORK • EVERYONE ELSE FORMS A THICK, DOUGHY LAYER OF INCOMPETENCE • RECIPE FOR A HEART ATTACK
LEFTOVER PIZZA TEAM	• EVERYONE HAS MOVED ON TO NEW TEAMS AND PROJECTS • EVERYONE BUT BOB • BOB WILL BE HERE FOREVER
BURNT PIZZA TEAM	• MISSED ANOTHER "DEADLINE" • LOOKS LIKE WE'RE WORKING ALL WEEKEND AGAIN • DOES ANYONE ELSE SMELL SMOKE?

@forrestbrazeal

Product requirements

- Business analysis
- Product requirements
- Development specs

Development iterations

Developers
QA engineers
(Two-pizza team)

Operations / Production

- Operations resiliency
- Infra engineering
- Data management

Thank You.