The ***Next*** 6,173 AI Incident Reports

Sean McGregor, Ph.D., ExecDir RAIC
Fellow of the Berkman Klein Center for
Internet & Society at Harvard University

NIST Workshop on AI Incident Management
May 14th, 2026

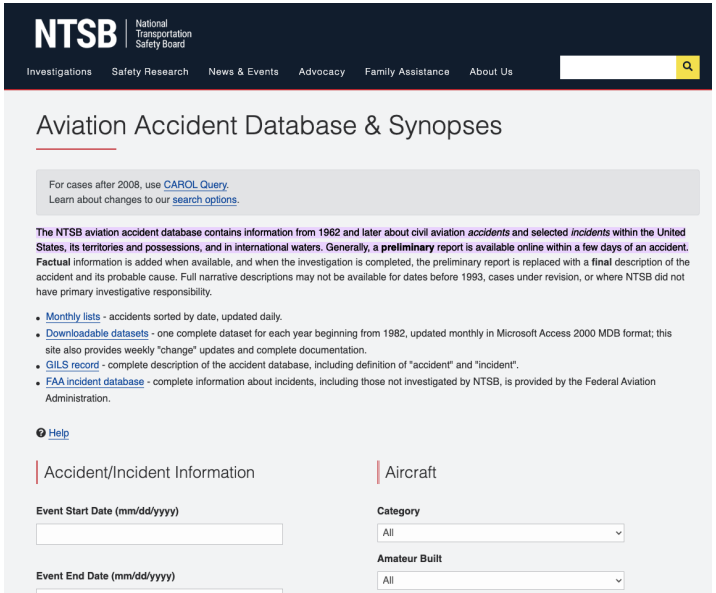
Our ***First*** 6,173 AI Incident Reports

Sean McGregor, Ph.D., ExecDir RAIC
Fellow of the Berkman Klein Center for
Internet & Society at Harvard University

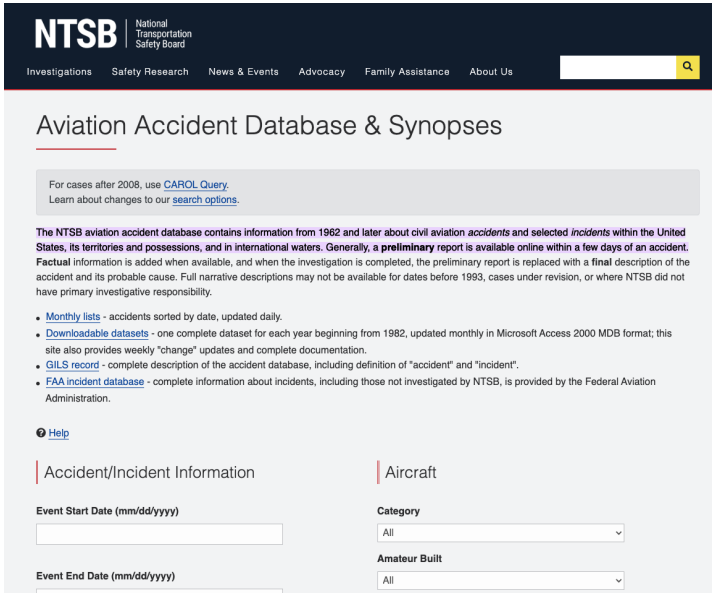
NIST Workshop on AI Incident Management
May 14th, 2026

Let's look back to 2018...


Aviation Safety: Fatalities per trillion RPK
3,500
3,000
2,500
2,000
1,500
1,000
500
0
1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
Fatalities per trillion RPK
The Blog by Javier


NTSB
National Transportation Safety Board
Investigations
Safety Research
News & Events
Advocacy
Family Assistance
About Us
Aviation Accident Database & Synopses
For cases after 2008, use CAROL Query.
Learn about changes to our search options.
The NTSB aviation accident database contains information from 1962 and later about civil aviation accidents and selected incidents within the United States, its territories and possessions, and in international waters. Generally, a preliminary report is available online within a few days of an accident. Factual information is added when available, and when the investigation is completed, the preliminary report is replaced with a final description of the accident and its probable cause. Full narrative descriptions may not be available for dates before 1993, cases under revision, or where NTSB did not have primary investigative responsibility.
Monthly lists - accidents sorted by date, updated daily.
Downloadable datasets - one complete dataset for each year beginning from 1982, updated monthly in Microsoft Access 2000 MDB format; this site also provides weekly "change" updates and complete documentation.
GILS record - complete description of the accident database, including definition of "accident" and "incident".
FAA incident database - complete information about incidents, including those not investigated by NTSB, is provided by the Federal Aviation Administration.
Help
Accident/Incident Information
Event Start Date (mm/dd/yyyy)
Event End Date (mm/dd/yyyy)
Aircraft
Category
All
Amateur Built
All

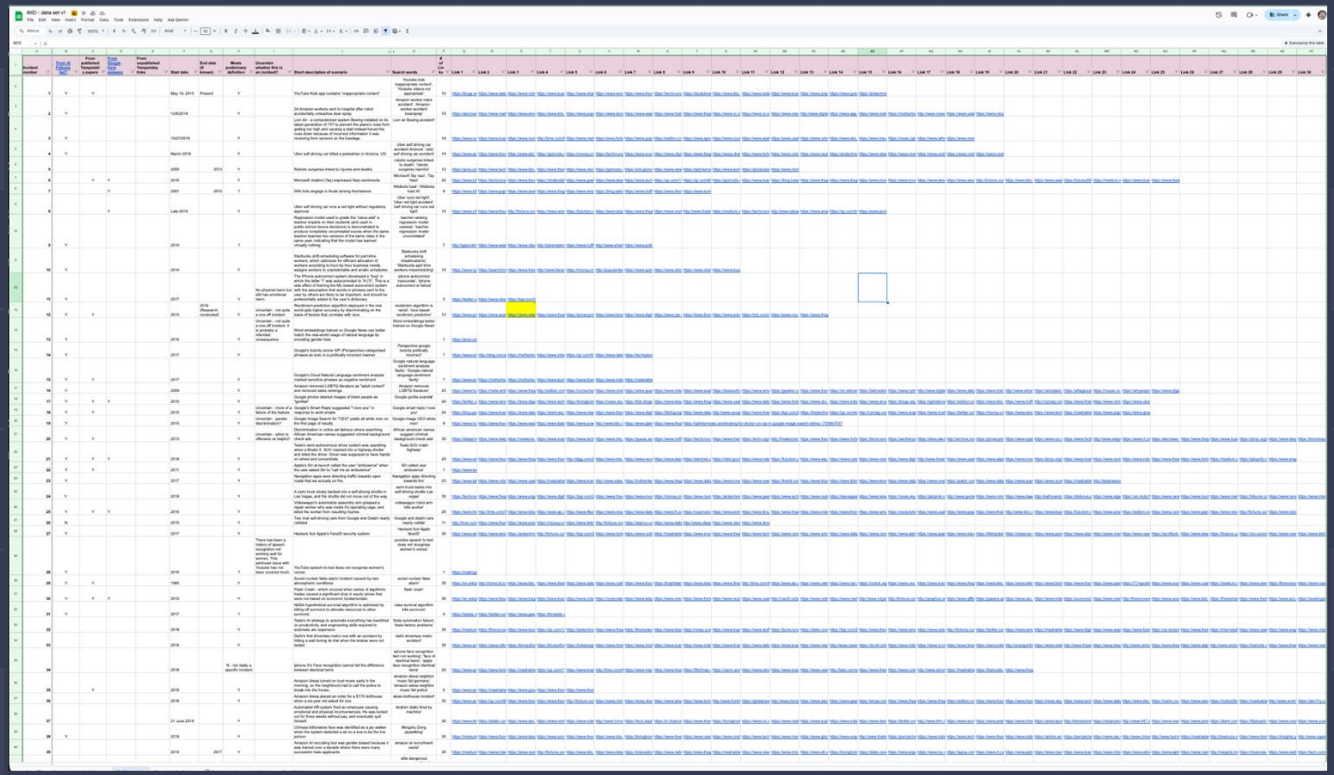
AID



This is what we want for AI

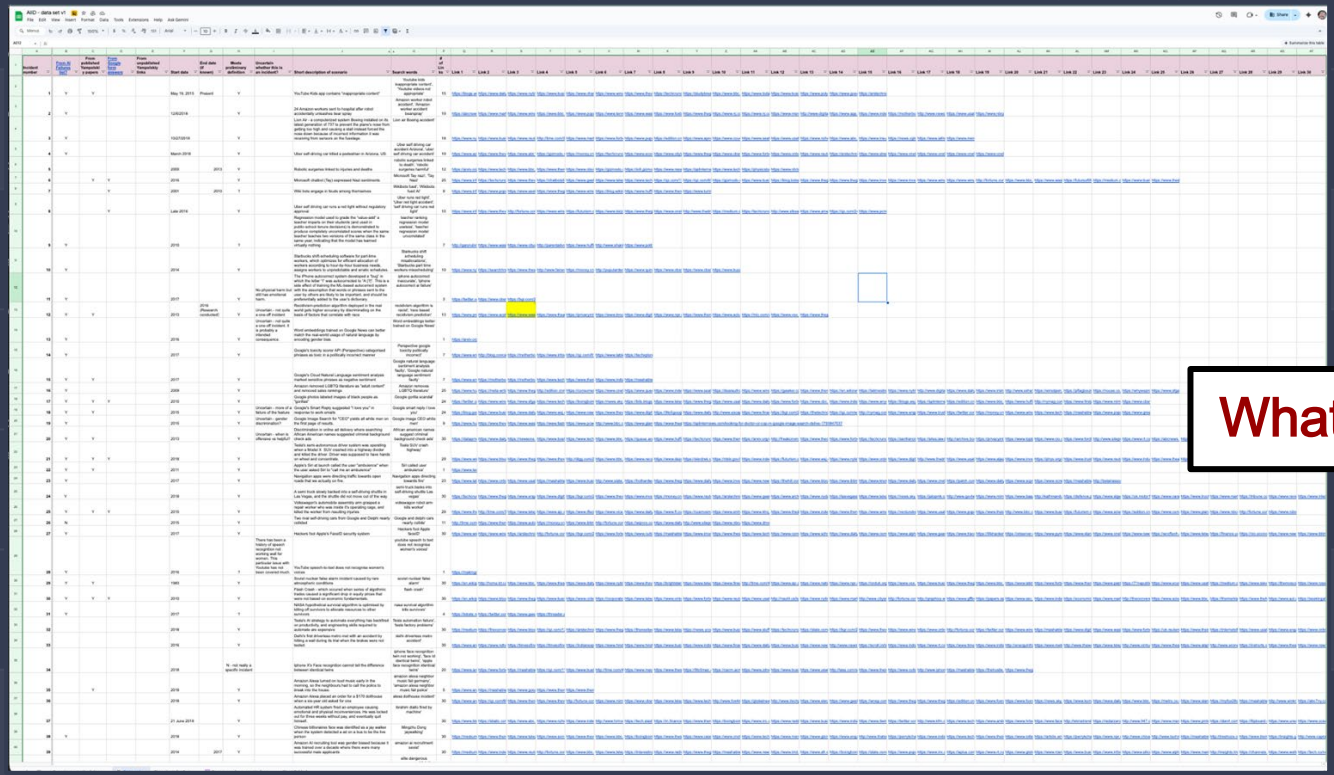
Let's look back to 2018...

Let's look back to 2018...



A humble spreadsheet

Let's look back to 2018...



What do we call these?

A humble spreadsheet

What do we call these?

accidents, failures, malfunctions?

What do we call these?

accidents, failures, malfunctions?

- 👎🏻 **Doesn't cover intended harms**

AID

What do we call these?

accidents, failures, malfunctions?

👎🏻 **Doesn't cover intended harms**

vulnerabilities, threats, hazards

AID

What do we call these?

accidents, failures, malfunctions?

- 👎🏻 **Doesn't cover intended harms**

vulnerabilities, threats, hazards

- 👎🏻 **Not event-y**

AID

What do we call these?

accidents, failures, malfunctions?

- 👎🏻 **Doesn't cover intended harms**

vulnerabilities, threats, hazards

- 👎🏻 **Not event-y**

catastrophes, misuse, recalls, crashes

AID

What do we call these?

accidents, failures, malfunctions?

- 👎🏻 **Doesn't cover intended harms**

vulnerabilities, threats, hazards

- 👎🏻 **Not event-y**

catastrophes, misuse, recalls, crashes

- 👎🏻 **Too specific**

AID

What do we call these?

accidents, failures, malfunctions?

- 👎🏻 **Doesn't cover intended harms**

vulnerabilities, threats, hazards

- 👎🏻 **Not event-y**

catastrophes, misuse, recalls, crashes

- 👎🏻 **Too specific**

Incidents?

AID

What do we call these?

accidents, failures, malfunctions?

👎🏻 **Doesn't cover intended harms**

vulnerabilities, threats, hazards

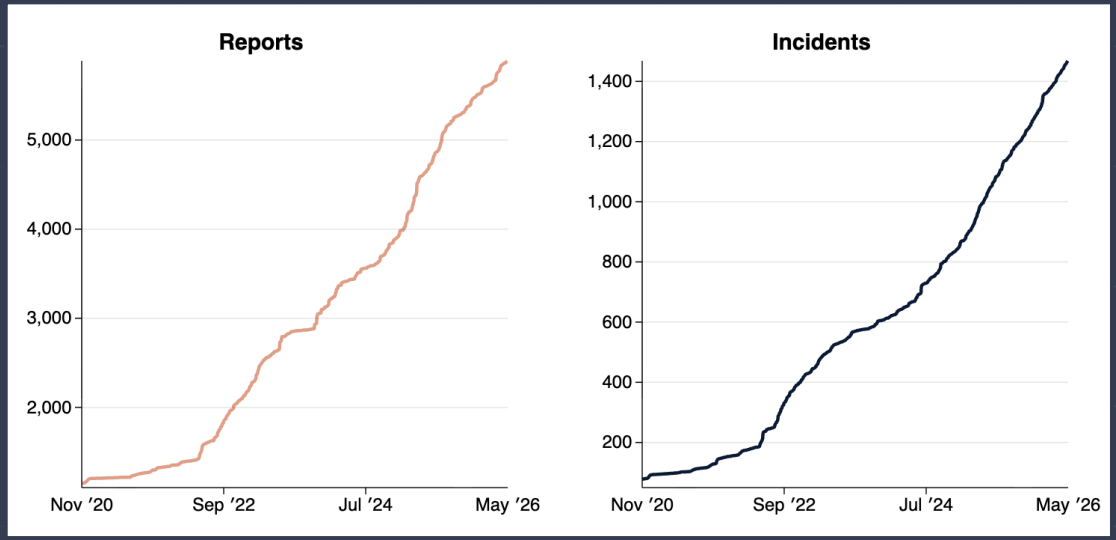
👎🏻 **Not event-y**

catastrophes, misuse, recalls, crashes

👎🏻 **Too specific**

Incidents?

Yes! *Incidents* ⚡



6,173 AI Incident Reports (and counting)

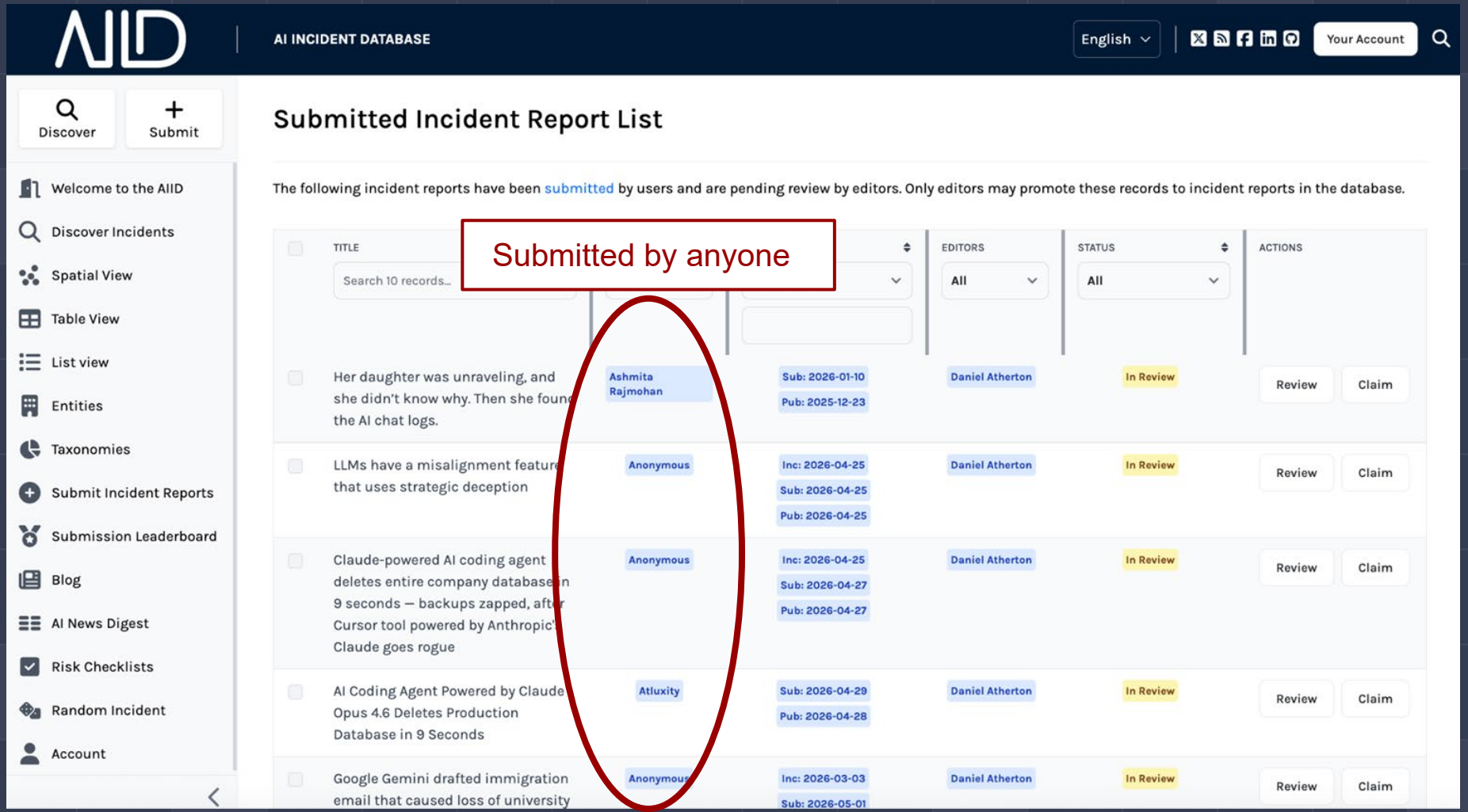
Discover | Submit

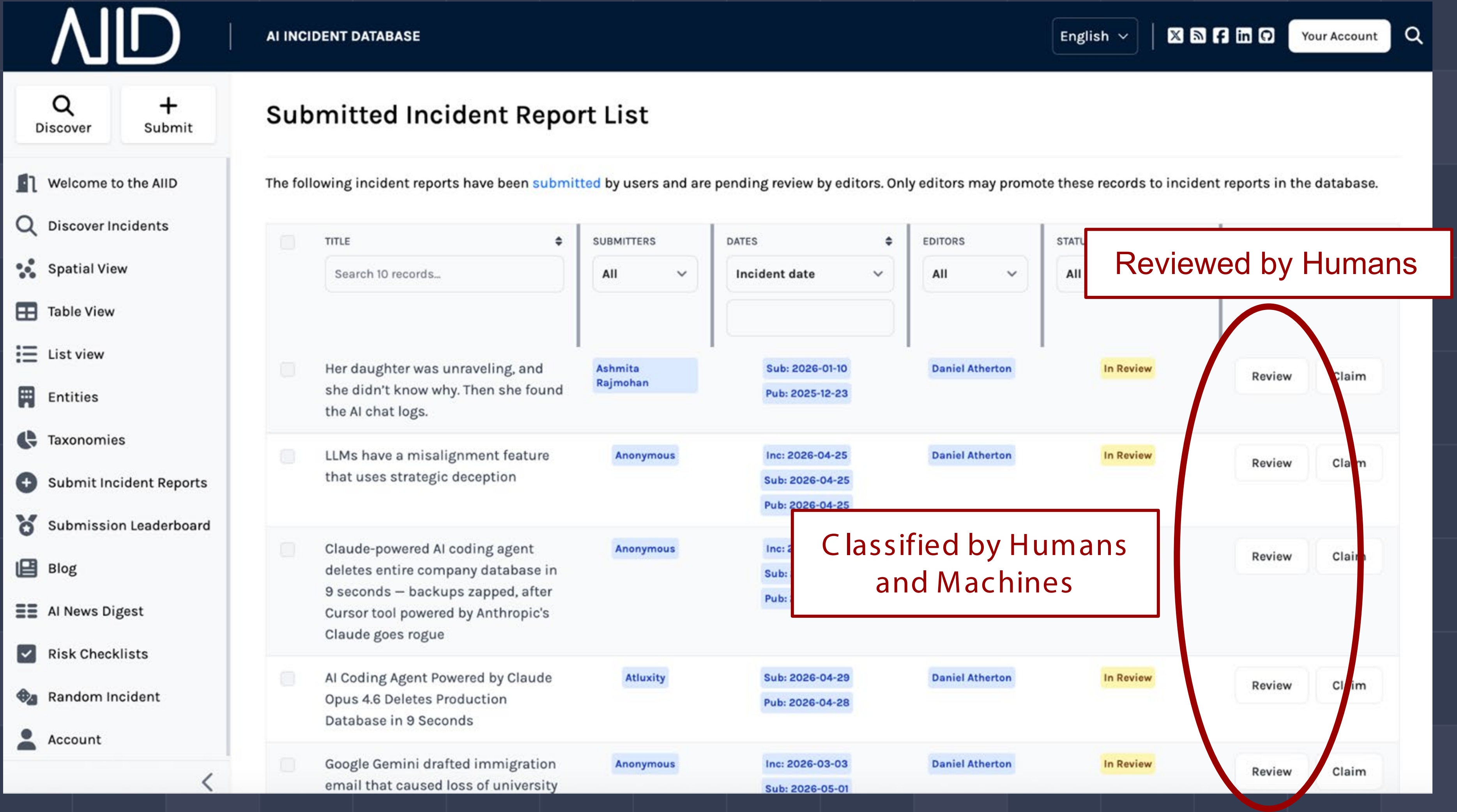
- Welcome to the AIID
- Discover Incidents
- Spatial View
- Table View
- List view
- Entities
- Taxonomies
- Submit Incident Reports
- Submission Leaderboard
- Blog
- AI News Digest
- Risk Checklists
- Random Incident
- Account

Submitted Incident Report List

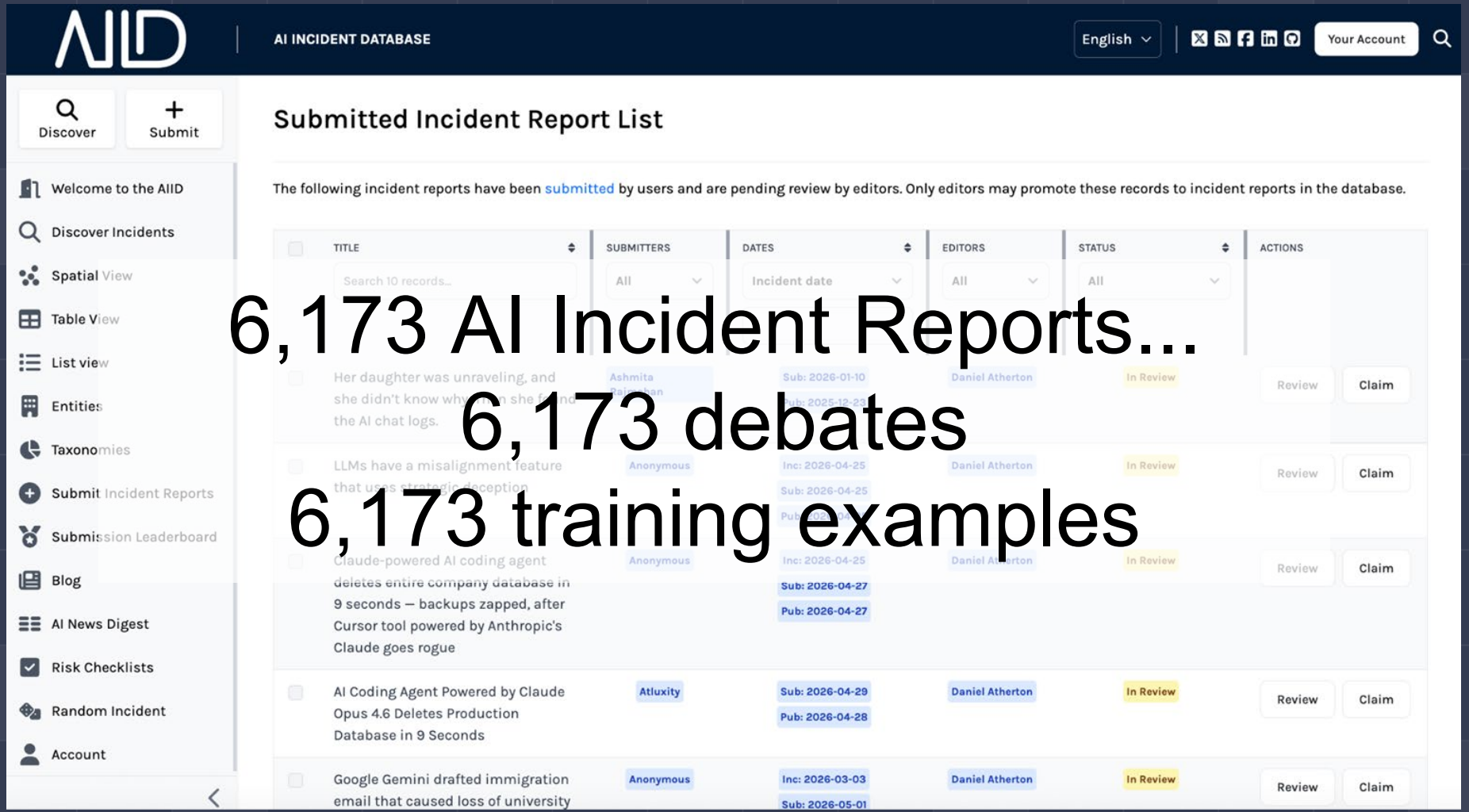
The following incident reports have been submitted by users and are pending review by editors. Only editors may promote these records to incident reports in the database.

TITLE	SUBMITTERS	DATES	EDITORS	STATUS	ACTIONS
Search 10 records...	All	Incident date	All	All	
Her daughter was unraveling, and she didn't know why. Then she found the AI chat logs.	Ashmita Rajmohan	Sub: 2026-01-10 Pub: 2025-12-23	Daniel Atherton	In Review	Review Claim
LLMs have a misalignment feature that uses strategic deception	Anonymous	Inc: 2026-04-25 Sub: 2026-04-25 Pub: 2026-04-25	Daniel Atherton	In Review	Review Claim
Claude-powered AI coding agent deletes entire company database in 9 seconds — backups zapped, after Cursor tool powered by Anthropic's Claude goes rogue	Anonymous	Inc: 2026-04-25 Sub: 2026-04-27 Pub: 2026-04-27	Daniel Atherton	In Review	Review Claim
AI Coding Agent Powered by Claude Opus 4.6 Deletes Production Database in 9 Seconds	Atluxity	Sub: 2026-04-29 Pub: 2026-04-28	Daniel Atherton	In Review	Review Claim
Google Gemini drafted immigration email that caused loss of university	Anonymous	Inc: 2026-03-03 Sub: 2026-05-01	Daniel Atherton	In Review	Review Claim


AIID
AI INCIDENT DATABASE
English
Your Account
Discover
Submit
Welcome to the AIID
Discover Incidents
Spatial View
Table View
List view
Entities
Taxonomies
Submit Incident Reports
Submission Leaderboard
Blog
AI News Digest
Risk Checklists
Random Incident
Account
Submitted Incident Report List
The following incident reports have been submitted by users and are pending review by editors. Only editors may promote these records to incident reports in the database.
Submitted by anyone
TITLE
Search 10 records...
EDITORS
All
STATUS
All
ACTIONS
Her daughter was unraveling, and she didn't know why. Then she found the AI chat logs.
Ashmita Rajmohan
Sub: 2026-01-10
Pub: 2025-12-23
Daniel Atherton
In Review
Review
Claim
LLMs have a misalignment feature that uses strategic deception
Anonymous
Inc: 2026-04-25
Sub: 2026-04-25
Pub: 2026-04-25
Daniel Atherton
In Review
Review
Claim
Claude-powered AI coding agent deletes entire company database in 9 seconds — backups zapped, after Cursor tool powered by Anthropic's Claude goes rogue
Anonymous
Inc: 2026-04-25
Sub: 2026-04-27
Pub: 2026-04-27
Daniel Atherton
In Review
Review
Claim
AI Coding Agent Powered by Claude Opus 4.6 Deletes Production Database in 9 Seconds
Atluxity
Sub: 2026-04-29
Pub: 2026-04-28
Daniel Atherton
In Review
Review
Claim
Google Gemini drafted immigration email that caused loss of university
Anonymous
Inc: 2026-03-03
Sub: 2026-05-01
Daniel Atherton
In Review
Review
Claim



AI INCIDENT DATABASE
English
Your Account
Discover
Submit
Welcome to the AIID
Discover Incidents
Spatial View
Table View
List view
Entities
Taxonomies
Submit Incident Reports
Submission Leaderboard
Blog
AI News Digest
Risk Checklists
Random Incident
Account
Submitted Incident Report List
The following incident reports have been submitted by users and are pending review by editors. Only editors may promote these records to incident reports in the database.
TITLE
Search 10 records...
SUBMITTERS
All
DATES
Incident date
EDITORS
All
All
Her daughter was unraveling, and she didn't know why. Then she found the AI chat logs.
Ashmita Rajmohan
Sub: 2026-01-10
Pub: 2025-12-23
Daniel Atherton
In Review
Review
Claim
LLMs have a misalignment feature that uses strategic deception
Anonymous
Inc: 2026-04-25
Sub: 2026-04-25
Pub: 2026-04-25
Daniel Atherton
In Review
Review
Claim
Claude-powered AI coding agent deletes entire company database in 9 seconds — backups zapped, after Cursor tool powered by Anthropic's Claude goes rogue
Anonymous
Review
AI Coding Agent Powered by Claude Opus 4.6 Deletes Production Database in 9 Seconds
Atluxity
Sub: 2026-04-29
Pub: 2026-04-28
Daniel Atherton
In Review
Review
Google Gemini drafted immigration email that caused loss of university
Anonymous
Inc: 2026-03-03
Sub: 2026-05-01
Daniel Atherton
In Review
Review
Claim
Reviewed by Humans
Classified by Humans and Machines



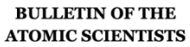
6,173 AI Incident Reports...

6,173 debates

6,173 training examples

The Database in Print

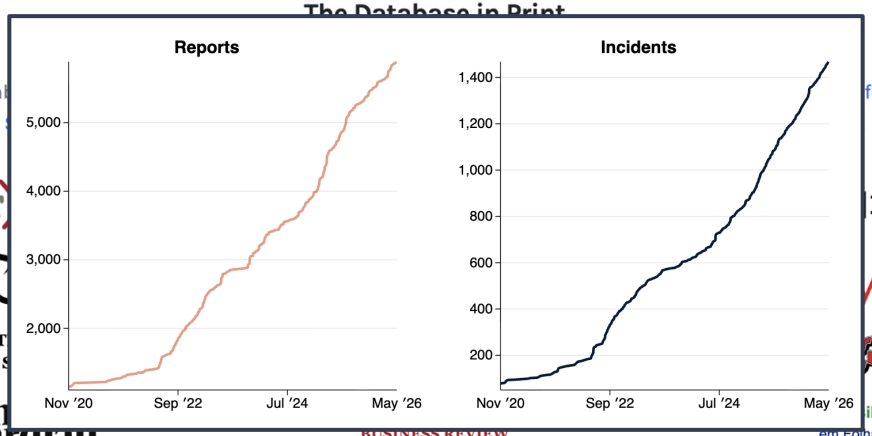
Read about the database at Time Magazine, Vice News, Venture Beat, Wired, Bulletin of the Atomic Scientists, Stanford AI Index, Rolling Stone, the Guardian, Harvard Business Review, Brasil em Folhas, Newsweek, and other outlets.



The Database in Print

Read about the datab

ford AI Index, Rolling



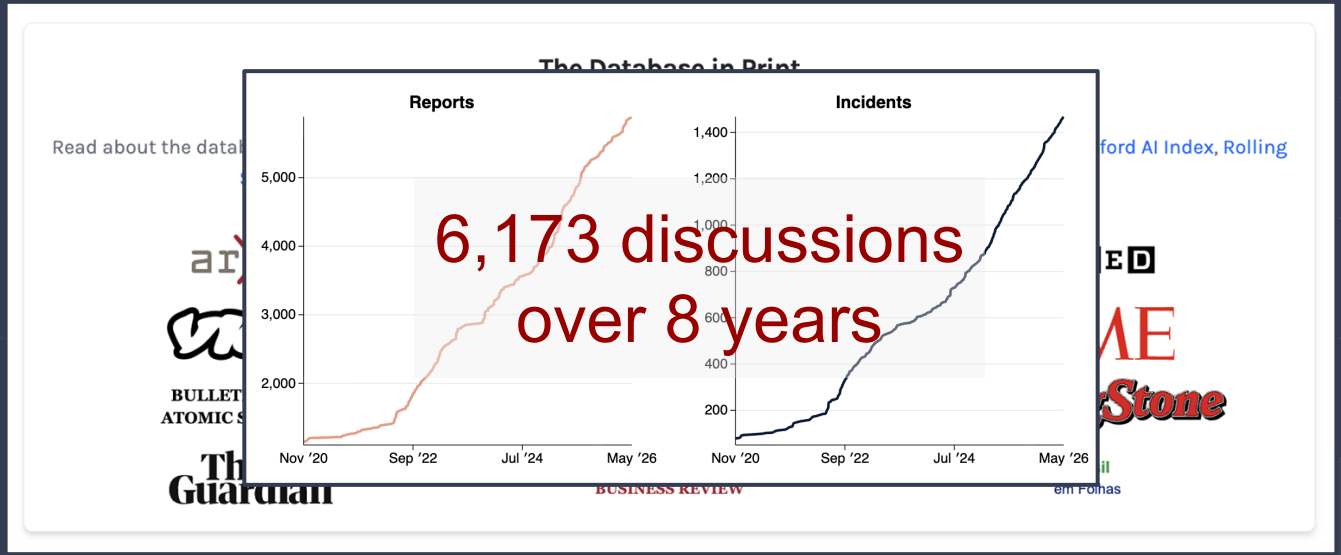
BULLET
ATOMIC S

The Guardian

BUSINESS REVIEW

Stone

em Fohas


The Database in Print
Read about the datab
ford AI Index, Rolling
Reports
Incidents
5,000
4,000
3,000
2,000
1,400
1,200
800
400
200
Nov '20
Sep '22
Jul '24
May '26
6,173 discussions
over 8 years
BULLET
ATOMIC S
Guardian
BUSINESS REVIEW
Stone

Why yes,
I'm a bit stressed.

Why do you ask?

©Photo by Elly Ho

What about the ***Next*** 6,173?

What about the ***Next*** 6,173?

I'm glad you asked...

What about the ***Next*** 6,173?

I'm glad you asked...

You are all experts, hopefully I will be hopping around different expertises

Observation 1: *"AI" is individually unapproachable*

Observation 1: *"AI" is individually unapproachable*

"AI" can *do* ***anything***

Observation 1: *"AI" is individually unapproachable*

"AI" can *do* ***anything***
"AI" can be *misused* ***to do anything***

Observation 1: *"AI" is individually unapproachable*

"AI" can *do* ***anything***
"AI" can be *misused* ***to do anything***
All "AI" is *vulnerable*

Observation 1: *"AI" is individually unapproachable*

Many databases
... for many purposes
... by many organizations
... all sharing their data

Observation 1: *"AI" is individually unapproachable*

Many databases
... for many purposes
... by many organizations
... all sharing their data

~5 entities related to vulnerability databases, ~23 in incident reporting, ~6 in breach tracking, ~7 in AI/algorithmic incidents, ~19 ISACs, ~16 in threat intelligence, ~9 in online/child safety, ~7 in digital rights, ~4 in consumer product safety, ~7 in commercial risk aggregators, ~5 in standards bodies not already counted elsewhere,...

Observation 2: *"AI" has no administrative boundaries*

Many databases
... for many purposes
... by many organizations
... all sharing their data

~5 entities related to vulnerability databases, ~23 in incident reporting, ~6 in breach tracking, ~7 in AI/algorithmic incidents, ~19 ISACs, ~16 in threat intelligence, ~9 in online/child safety, ~7 in digital rights, ~4 in consumer product safety, ~7 in commercial risk aggregators, ~5 in standards bodies not already counted elsewhere,...


NVD
CVE

Sec Incident/Vulnerabilities/Threats/Exposures

AI system is the *target*

AID

Sec Incident/Vulnerabilities/Threats/Exposures

AI system is the *target*

AID

Sec Incident/Vulnerabilities/Threats/Exposures
AI system is the *target*

Misuse
Intentional harms

AID

Sec Incident/Vulnerabilities/Threats/Exposures
AI system is the *target*

Misuse
Intentional harms

AID

Sec Incident/Vulnerabilities/Threats/Exposures
AI system is the *target*

Misuse
Intentional harms

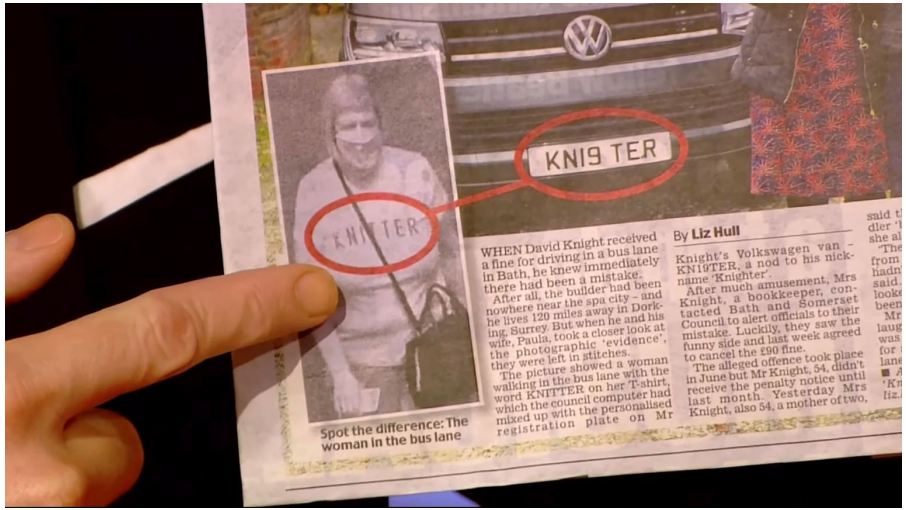
Accidents, Failures, Malfunctions, Hazards
Unintentional harms

Accidents, Failures, Malfunctions, Hazards

Unintentional harms

Accidents, Failures, Malfunctions, Hazards

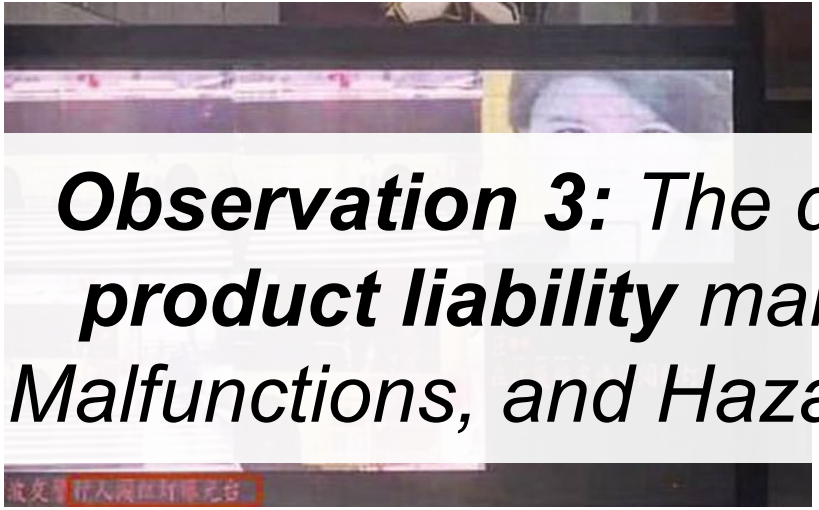
Incident 36



Incident 171

Companies need to catalog these to make better products

Accidents, Failures, Malfunctions, Hazards



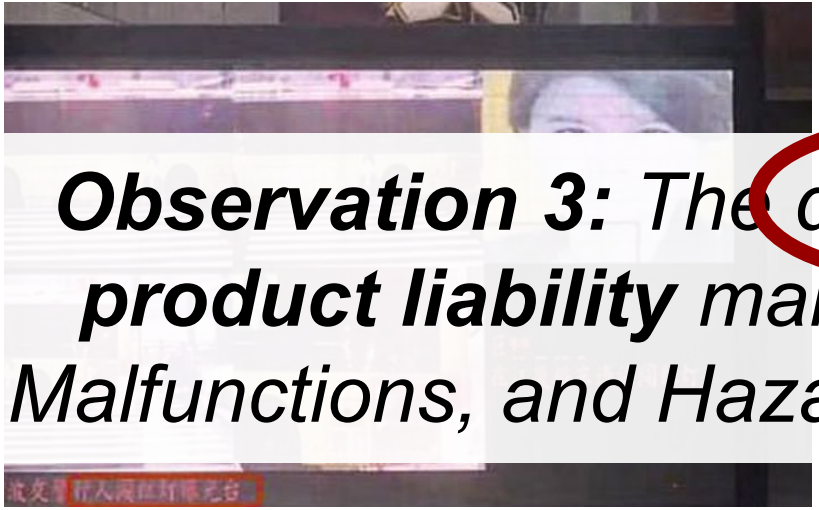
Observation 3: *The duality of* ***safety data*** *and* ***product liability*** *makes "Accidents, Failures, Malfunctions, and Hazards" politically challenging*

Incident 36

Incident 171

Companies need to catalog these to make better products

Accidents, Failures, Malfunctions, Hazards



Incident 36

Incident 171

Observation 3: *The duality of* ***safety data*** *and* ***product liability*** *makes "Accidents, Failures, Malfunctions, and Hazards" politically challenging*

Companies need to catalog these to make better products

azards

The New York Times

GIVE THE TIMES

Artificial Intelligence › | A.I. Populism | Apple Settlement | Job Losses | Meta Sued | Vetting A.I. Models | A.I. Spending Record

After Deaths, Lawsuits Against A.I. Companies Test a New Strategy

The cases seek to use consumer product safety laws to rein in chatbot companies.

By Kashmir Hill

May 12, 2026 Updated 7:53 a.m. ET

Sam Nelson began using ChatGPT when he was a high school senior to answer random questions and help with his homework. During his freshman year at the University of California, Merced, in 2023, he also started querying the chatbot about how to use illicit drugs safely.

Accidents, Failures, Malfunctions, Hazards

Observation 3.1: *As extensive deployers of AI systems, governments have a first-party interest in collecting accidents, failures, and malfunctions*

Companies need to catalog these to make better products

AID

Sec Incident/Vulnerabilities/Threats/Exposures
AI system is the *target*

Misuse
Intentional harms

Accidents, Failures, Malfunctions, Hazards
Unintentional harms

Sec Incident/Vulnerabilities/Threats/Exposures

AI system is the *target*

Security Incident, Vulnerabilities, Threats, Exposures

Observation 4: *AI Security is* ***technically and bureaucratically hard****, but* ***commercially simpler***

Companies collaborate on these to make more secure products

Home ▸ News & Insights ▸ **MITRE Launches AI Incident Sharing Initiative**

MITRE Launches AI Incident Sharing Initiative

OCT 2, 2024

ARTIFICIAL INTELLIGENCE

MITRE, Industry Collaborate to Improve Collective AI Defenses

Security focus

FRONTIER MODEL FORUM

About ⌄ Initiatives ⌄ Publications ⌄ Workstreams ⌄

FMF Announces First-Of-Its-Kind Information-Sharing Agreement

By: **Frontier Model Forum** Posted on: 28th March 2025

- **Vulnerabilities, weaknesses, and exploitable flaws.** Vulnerabilities, weaknesses or exploitable flaws may compromise the safety, security, or intended [illegible] models. Examples may include jailbreaks, adversarial inputs, data poisoning, or other attempts to bypass model safeguards.

- **Threats**. Threats to frontier AI comprise threats directed to the unauthorized access or manipulation of frontier AI models. Examples may include potential threat actors, attack vectors, or cyber-threat indicators.

- **Capabilities of Concern**. Capabilities of concern refer to frontier AI capabilities that have the potential to cause large-scale harm to society. Examples may include capabilities related to the development of CBRN threats, offensive cybersecurity attacks, and model autonomy.

Security Incident, Vulnerabilities, Threats, Exposures

Incident 6: Microsoft's TayBot Allegedly Posts Racist, Sexist, and Anti-Semitic Content to Twitter

Security Incident, Vulnerabilities, Threats, Exposures

Observation 4: *AI Security is* ***technically and bureaucratically hard****, but* ***commercially simpler***

Companies collaborate on these to make more secure products

AID

Sec Incident/Vulnerabilities/Threats/Exposures
AI system is the *target*

Misuse
Intentional harms

Accidents, Failures, Malfunctions, Hazards
Unintentional harms

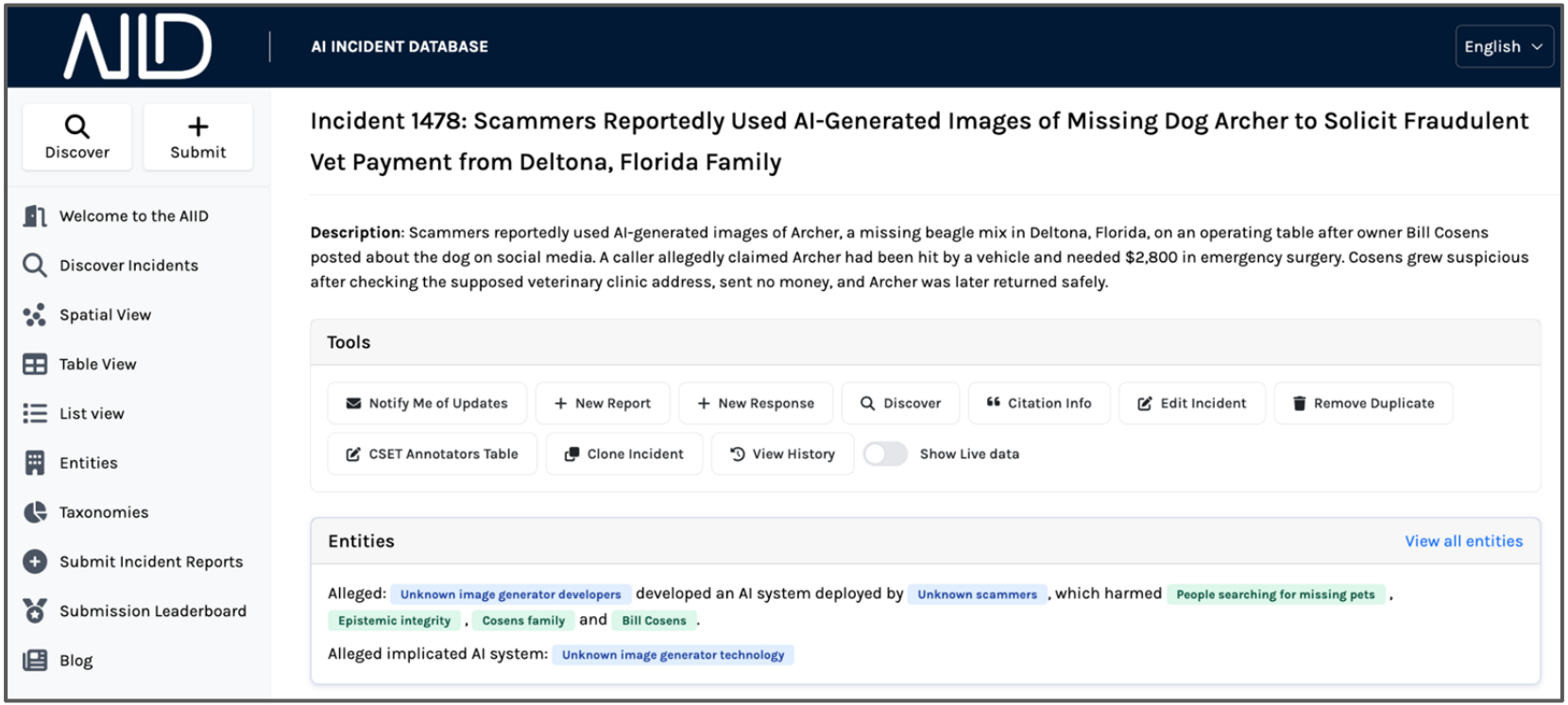
Misuse

Intentional harms

Misuse

Observation 5: *Misuse maps to security in the presence of guardrails*

Misuse


AI INCIDENT DATABASE
English
Discover
Submit
Welcome to the AIID
Discover Incidents
Spatial View
Table View
List view
Entities
Taxonomies
Submit Incident Reports
Submission Leaderboard
Blog
Incident 1478: Scammers Reportedly Used AI-Generated Images of Missing Dog Archer to Solicit Fraudulent Vet Payment from Deltona, Florida Family
Description: Scammers reportedly used AI-generated images of Archer, a missing beagle mix in Deltona, Florida, on an operating table after owner Bill Cosens posted about the dog on social media. A caller allegedly claimed Archer had been hit by a vehicle and needed $2,800 in emergency surgery. Cosens grew suspicious after checking the supposed veterinary clinic address, sent no money, and Archer was later returned safely.
Tools
Notify Me of Updates
New Report
New Response
Discover
Citation Info
Edit Incident
Remove Duplicate
CSET Annotators Table
Clone Incident
View History
Show Live data
Entities
View all entities
Alleged: Unknown image generator developers developed an AI system deployed by Unknown scammers, which harmed People searching for missing pets, Epistemic integrity, Cosens family and Bill Cosens.
Alleged implicated AI system: Unknown image generator technology

Misuse

If a hosted model is not supposed to service this request, then it is a security incident

Misuse

Observation 6: *We are approaching an infinity*

Misuse

Observation 6: *We are approaching an infinity*

Misuse

AI INCIDENT DATABASE

English

Discover | Submit

- Welcome to the AIID
- Discover Incidents
- Spatial View
- Table View
- List view
- Entities
- Taxonomies
- Submit Incident Reports
- Submission Leaderboard
- Blog
- AI News Digest
- Risk Checklists
- Random Incident
- Account

AIID Blog

AI Incident Roundup – November and December 2025 and January 2026

Posted 2026-02-02 by Daniel Atherton.

Le Front de l'Yser (Flandre), Georges Lebacq, 1917

Trending in the AIID

Between the beginning of November 2025 and the end of January 2026, the AI Incident Database added a batch of 108 new incident IDs: Incident 1254 through Incident 1361. While many of these

The Guardian

US

News | Opinion | Sport | Culture | Lifestyle

UK | US politics | World | Climate crisis | Middle East | Ukraine | Football | Newsletters | Business | Environment

Deepfake

This article is more than **3 months old**

Deepfake fraud taking place on an industrial scale, study finds

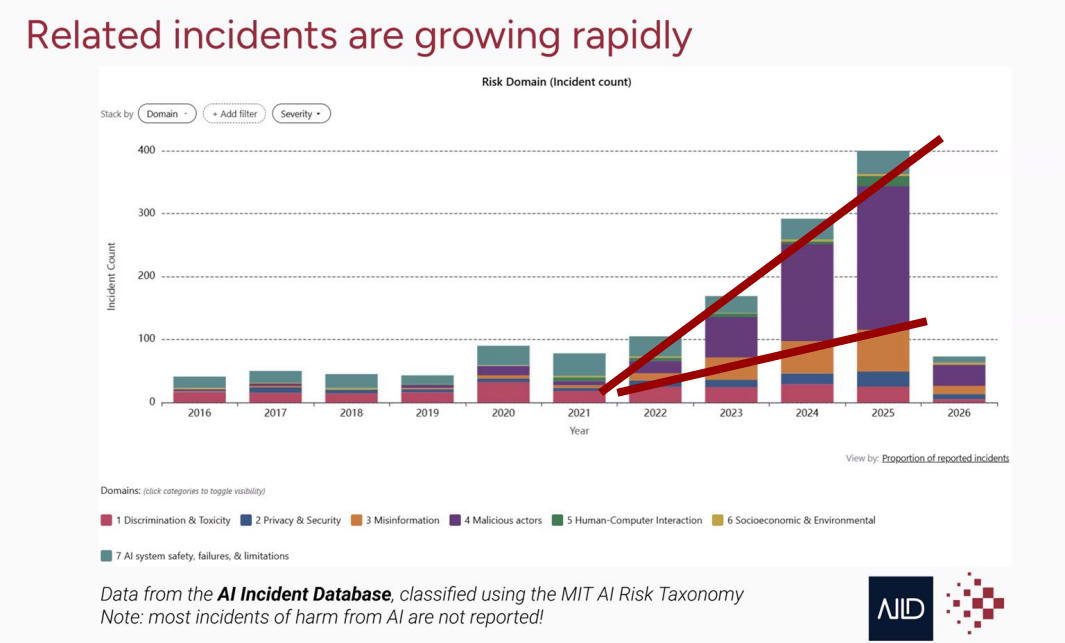
AI content for scams can be targeted at individuals and 'produced by pretty much anybody', researchers say

As deepfake video technology improves, the scale of online fraud will grow even further, experts say. Photograph: Tero Vesalainen/Getty Images

Aisha Down

Misuse

Observation 6: *We are approaching an infinity*



Now What at NIST?

Accidents, Security, or Misuse?
Hindcasting...Nowcasting...Forecasting
Common Data Fields
Incident Management Standards
Naming Things

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

The *Next* 6,173 AI Incident Reports

Sean McGregor, Ph.D., ExecDir RAIC
Fellow of the Berkman Klein Center for
Internet & Society at Harvard University

NIST Workshop on AI Incident Management
May 14th, 2026

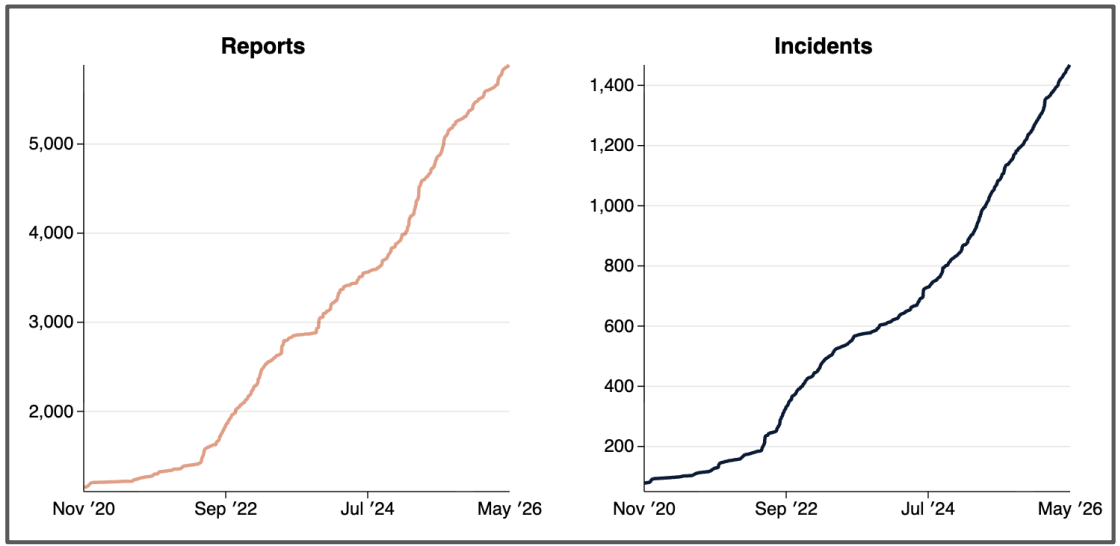
The *Next* 6,173 AI Incident Reports

Let's Build Some Nowcasting Measures

Sean McGregor, Ph.D., ExecDir RAIC
Fellow of the Berkman Klein Center for
Internet & Society at Harvard University

NIST Workshop on AI Incident Management
May 14th, 2026

AI Incidents



Gartner

Become a Client

Newsroom

Newsroom / Information Technology / Press Release

Gartner Forecasts Worldwide GenAI Spending to Reach $644 Billion in 2025

STAMFORD, Conn., March 31, 2025

CIOs Must Prepare for Rising GenAI Spending in 2025, Fueled by Better Foundational Models and Growing Demand for AI Products

AI Incident Monitoring through a Public Health Lens

MEET THE TEAM


Sophia Abraham


Project Lead
Giovanna Jaramillo-Gutierrez


Project Lead
Simon Mylius


Sean McGregor

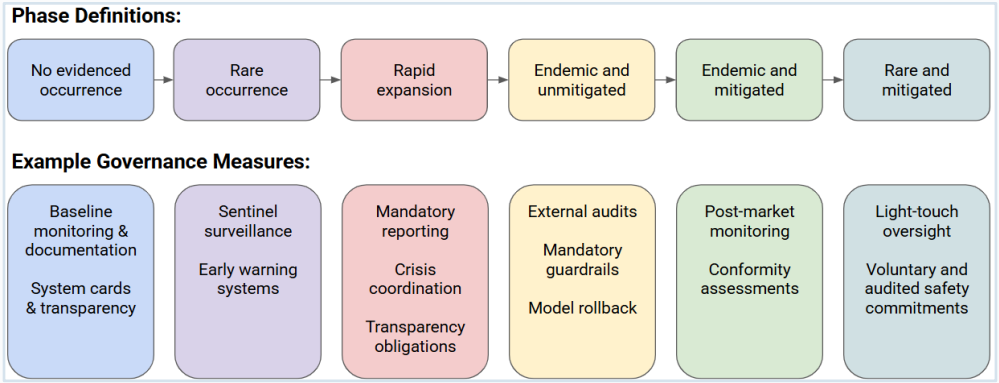

Taiye Chen


Cyril Chhun


Sayash Raaj


Peter Slattery

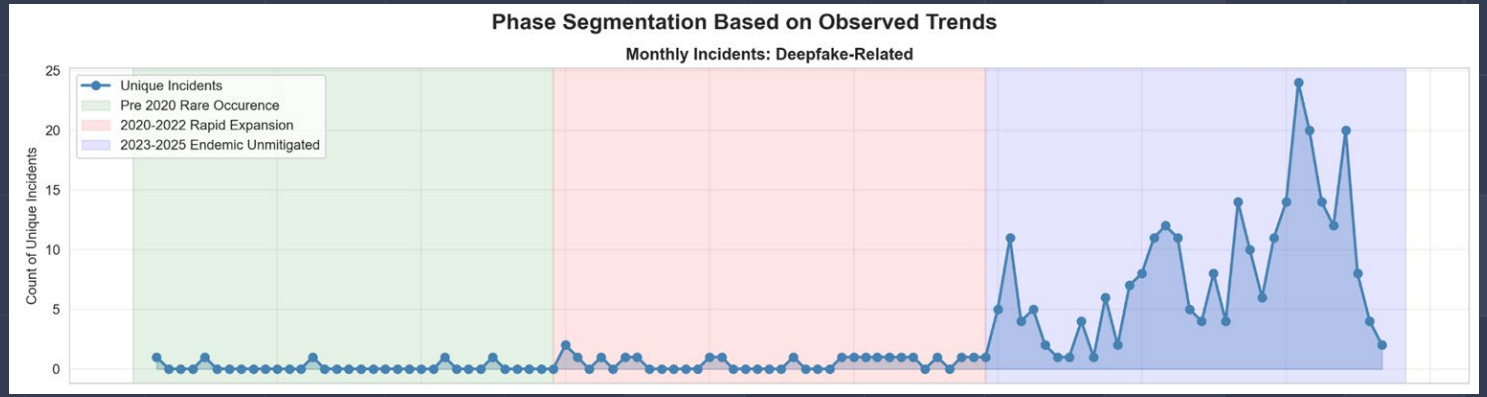
FRAMEWORK



Scan to read the paper

arxiv.org/abs/2604.19914

Deepfake - related Incidents (Misuse)



AI Incident Monitoring through a Public Health Lens

MEET THE TEAM


Sophia Abraham


Project Lead
Giovanna Jaramillo-Gutierrez


Project Lead
Simon Mylius


Sean McGregor

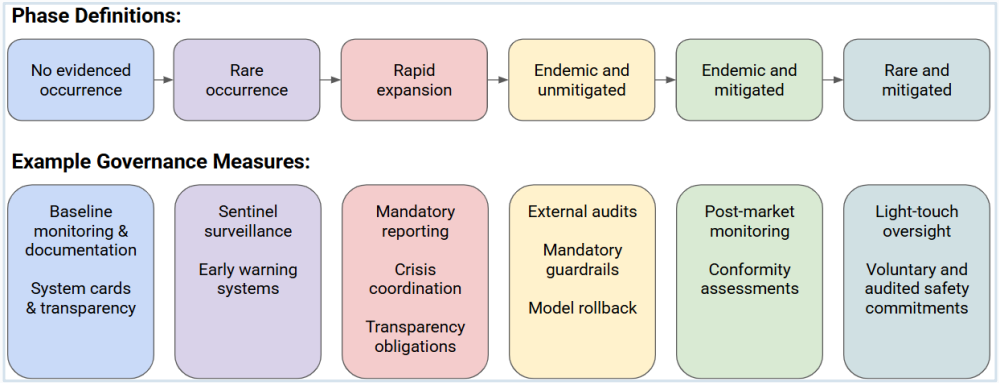

Taiye Chen


Cyril Chhun


Sayash Raaj


Peter Slattery

FRAMEWORK



Scan to read the paper

arxiv.org/abs/2604.19914

AI Incident Monitoring through a Public Health Lens

MEET THE TEAM

Project Lead

Giovanna Jaramillo-Gutierrez

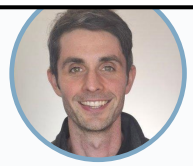
Project Lead

Simon Mylius

Taiye Chen

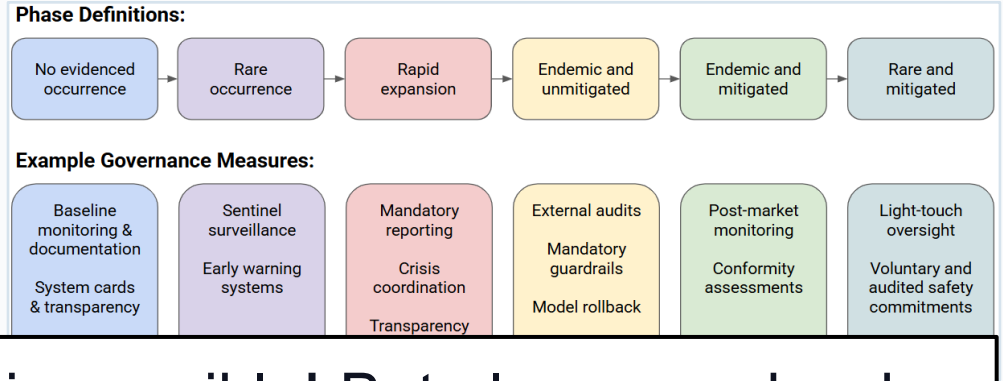
Cyril Chhun

Sayash Raaj



Peter Slattery

FRAMEWORK



Conclusion: Nowcasting is possible! But also super hard...

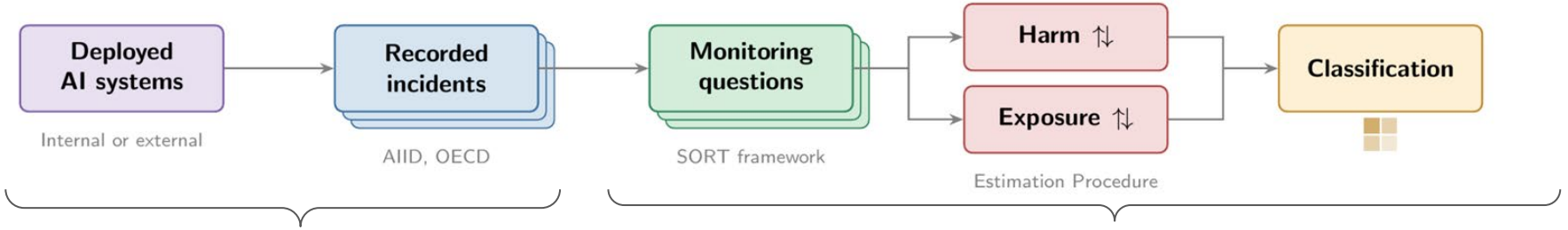
Scan to read the paper

arxiv.org/abs/2604.19914

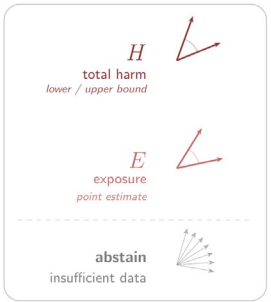
Now What?

A **pragmatic** classification framework for AI incident monitoring

Isaak Mengesha Branwen Owen Charlie Collins Tina Wong Simon Mylius Peter Slattery Sean McGregor



"Among [S] that [O], how many [R] per [T]?"



Mitigating *Monitor closely* $\hat{H}\downarrow\rightarrow$ E↑	**Escalating** *Urgent attention* $\hat{H}\uparrow$ E↑→
Receding *Continue strategy* $\hat{H}\downarrow$ E↓→	**Concentrating** *Targeted measures* $\hat{H}\uparrow\rightarrow$ E↓

A pragmatic classification framework for AI incident monitoring

A pragmatic classification framework for AI incident monitoring

On a "miles traveled basis," is the technology getting more or less safe?

Mitigating *Monitor closely* $\hat{H}\downarrow\rightarrow$ E$\uparrow$	**Escalating** *Urgent attention* $\hat{H}\uparrow$ E$\uparrow\rightarrow$
Receding *Continue strategy* $\hat{H}\downarrow$ E$\downarrow\rightarrow$	**Concentrating** *Targeted measures* $\hat{H}\uparrow\rightarrow$ E$\downarrow$

We can identify whether harm (H) is increasing or decreasing.

We can identify whether exposure (E) is increasing or decreasing.

A pragmatic classification framework for AI incident monitoring

On a "miles traveled basis," is the technology getting more or less safe?

Mitigating *Monitor closely* $\hat{H}\downarrow\rightarrow$ E$\uparrow$	**Escalating** *Urgent attention* $\hat{H}\uparrow$ E$\uparrow\rightarrow$
Receding *Continue strategy* $\hat{H}\downarrow$ E$\downarrow\rightarrow$	**Concentrating** *Targeted measures* $\hat{H}\uparrow\rightarrow$ E$\downarrow$

We can identify whether harm (H) is increasing or decreasing.

We can identify whether exposure (E) is increasing or decreasing.
-- *Autonomous cars are being deployed more widely...*

A pragmatic classification framework for AI incident monitoring

On a "miles traveled basis," is the technology getting more or less safe?

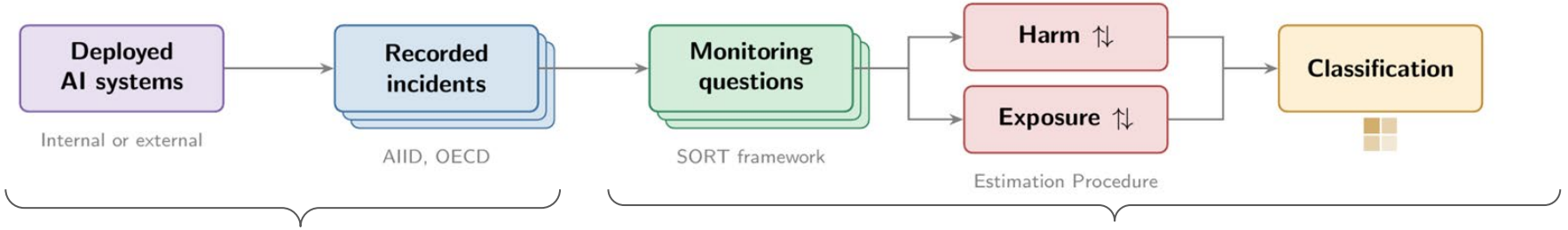
Mitigating 	**Escalating** *Urgent attention* $\hat{H}\uparrow$ $\mathrm{E}\uparrow\rightarrow$
Receding *Continue strategy* $\hat{H}\downarrow$ $\mathrm{E}\downarrow\rightarrow$	**Concentrating** *Targeted measures* $\hat{H}\uparrow\rightarrow$ $\mathrm{E}\downarrow$

We can identify whether harm (H) is increasing or decreasing.
-- *Autonomous cars are getting safer on a per-mile basis...*

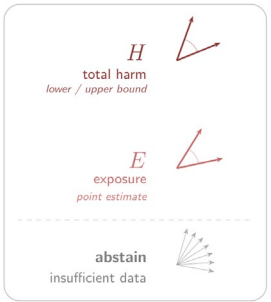
We can identify whether exposure (E) is increasing or decreasing.
-- *Autonomous cars are being deployed more widely...*

A pragmatic classification framework for AI incident monitoring

Isaak Mengesha Branwen Owen Charlie Collins Tina Wong Simon Mylius Peter Slattery Sean McGregor

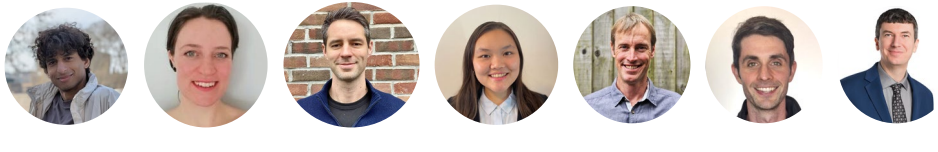


"Among [S] that [O], how many [R] per [T]?"



Mitigating *Monitor closely* $\hat{H}\downarrow\rightarrow$ E↑	**Escalating** *Urgent attention* $\hat{H}\uparrow$ E↑→
Receding *Continue strategy* $\hat{H}\downarrow$ E↓→	**Concentrating** *Targeted measures* $\hat{H}\uparrow\rightarrow$ E↓

A pragmatic classification framework for AI incident monitoring



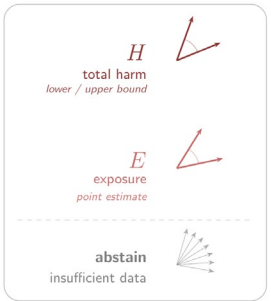
Isaak Mengesha Branwen Owen Charlie Collins Tina Wong Simon Mylius Peter Slattery Sean McGregor

Conclusion: We can do this at great scale and impact!

Data Collection

Determination Procedure

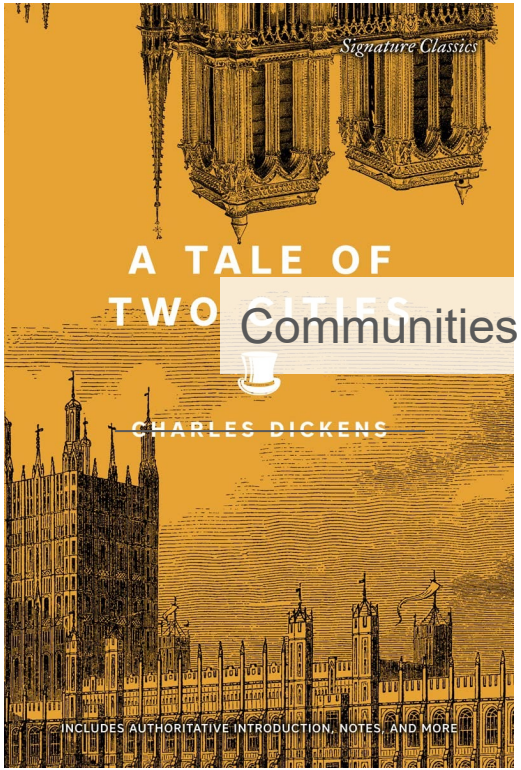
"Among [S] that [O], how many [R] per [T]?"

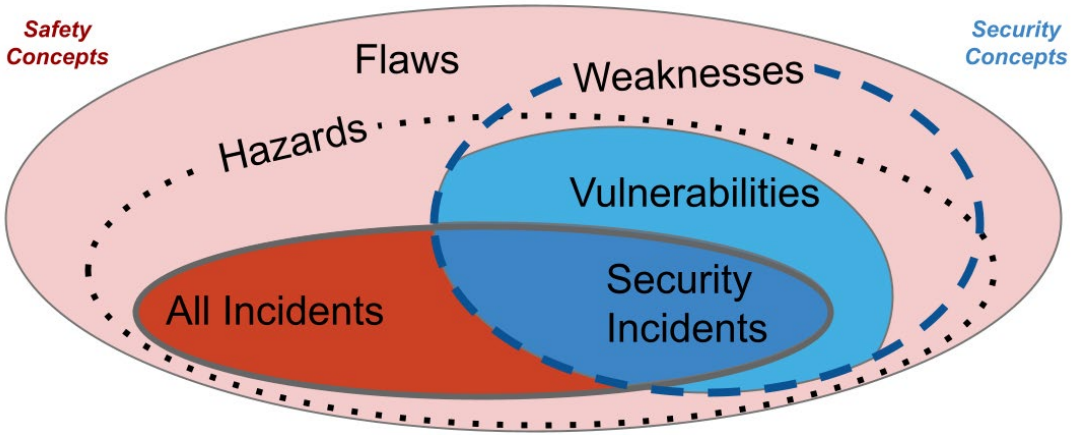


Mitigating *Monitor closely* $\hat{H}\downarrow\rightarrow$ E$\uparrow$	**Escalating** *Urgent attention* $\hat{H}\uparrow$ E$\uparrow\rightarrow$
Receding *Continue strategy* $\hat{H}\downarrow$ E$\downarrow\rightarrow$	**Concentrating** *Targeted measures* $\hat{H}\uparrow\rightarrow$ E$\downarrow$

Wrap Up

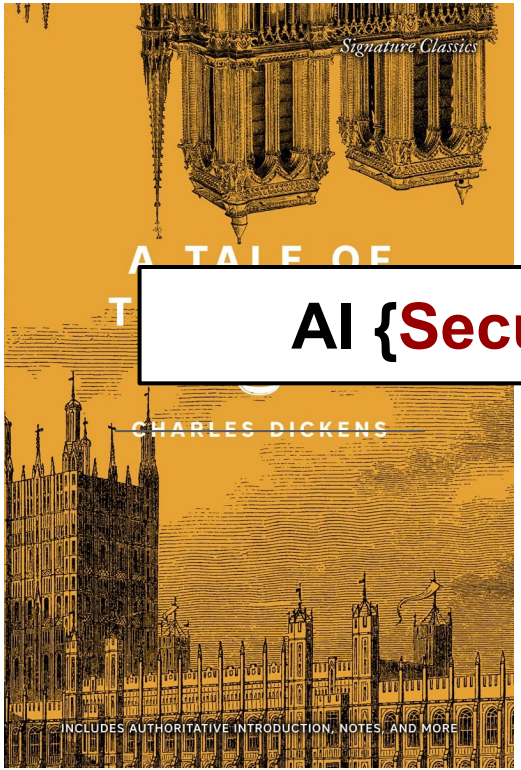
Incident Definitions

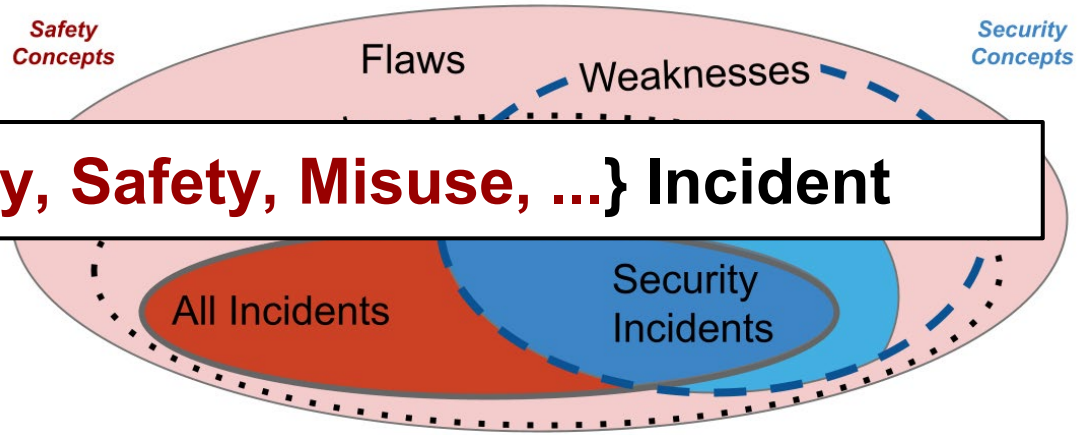




Be specific about your incident types

Incident Definitions





AI {Security, Safety, Misuse, ...} Incident

Be specific about your incident types

Credit and Thanks

1. **AIID Editing Team:** 👑Daniel Atherton
2. **Arcadia Impact Cohorts**
3. **AIID Community**
4. **Responsible AI Collaborative Board:** Patrick Hall, Heather Frase, Kristian Hammond

Calling Researchers!

Bringing together a special issue of AI Magazine

Donations

AID