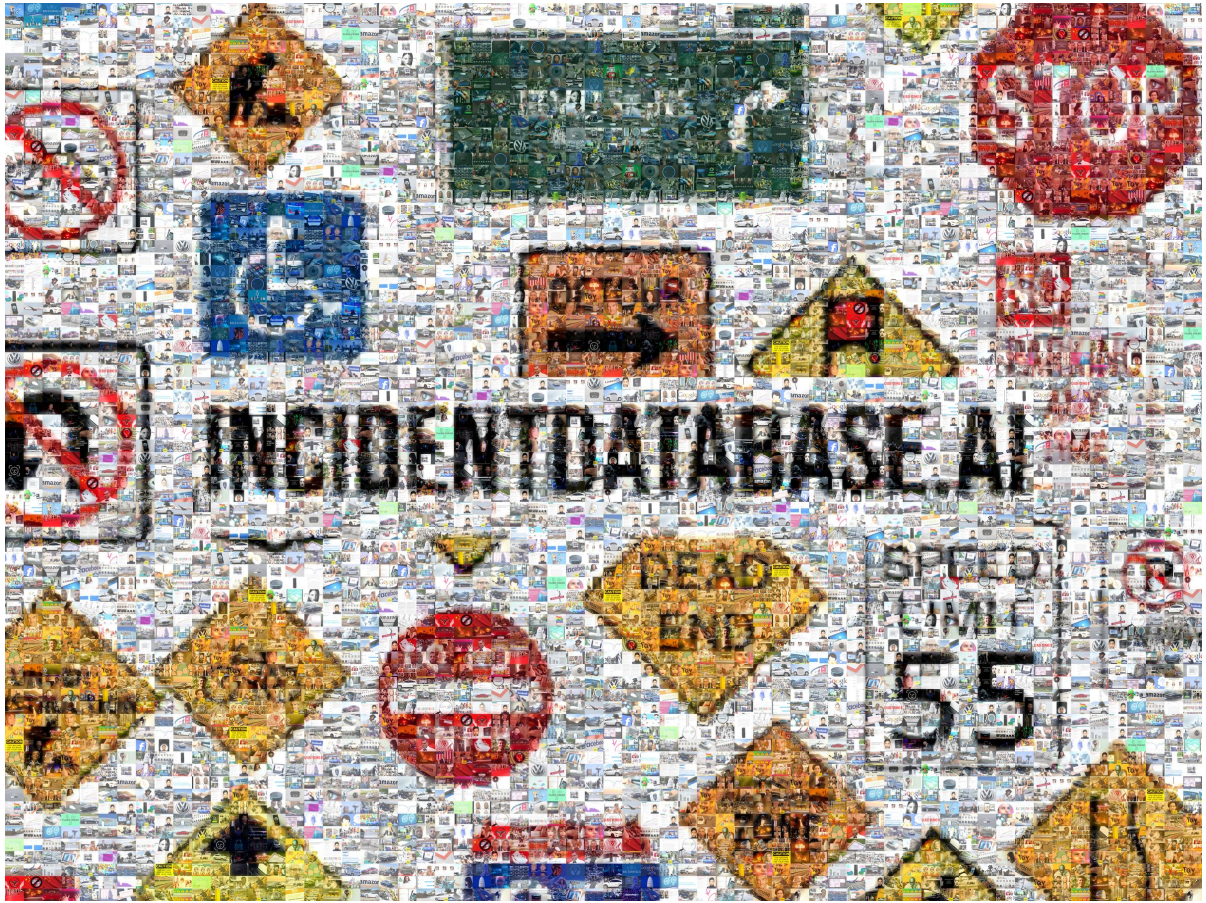


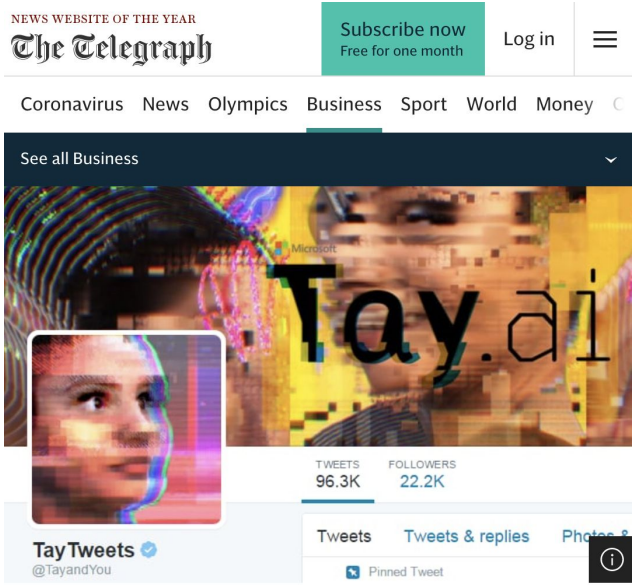
Preventing Repeated AI Harms by Sharing AI Failures

HCII 2021

Sean McGregor, Ph.D.

2021-07-26

PARTNERSHIP ON AI



Microsoft deletes 'teen girl' AI after it became a Hitler-loving sex robot within 24 hours

By Helena Horton

24 March 2016 • 3:37pm

A day after Microsoft introduced an innocent Artificial Intelligence chat robot to Twitter it has had to delete it after it transformed into an evil Hitler-loving, incestual sex-promoting, 'Bush did 9/11'-proclaiming robot. ...

March 2016

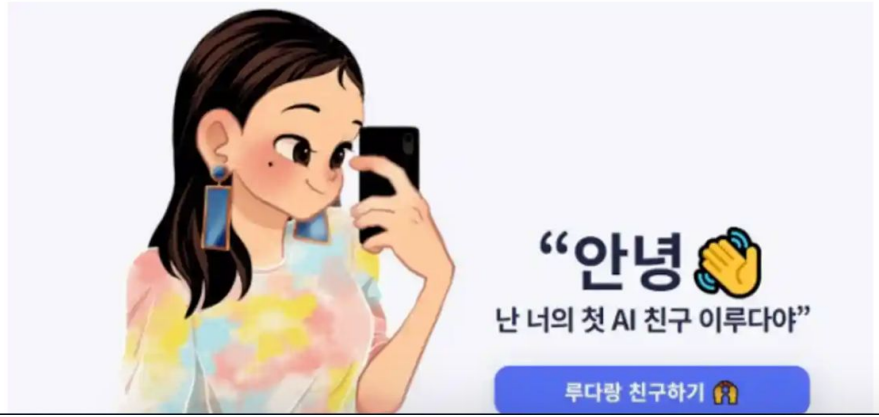
South Korea

South Korean AI chatbot pulled from Facebook after hate speech towards minorities

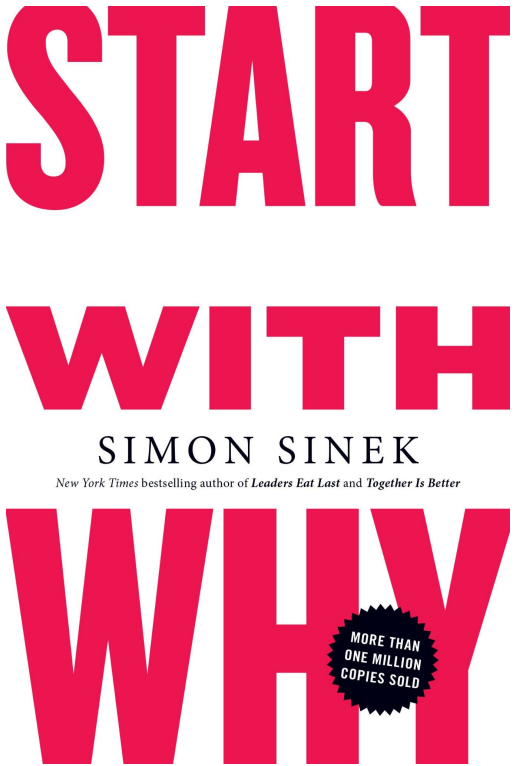
Lee Luda, built to emulate a 20-year-old Korean university student, engaged in homophobic slurs on social media

Justin McCurry *in Tokyo*

Wed 13 Jan 2021 23.24 EST



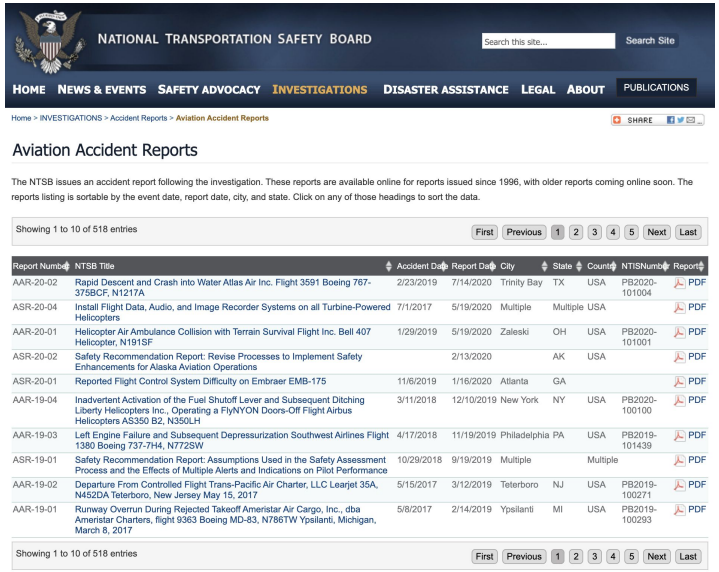
January 2021



"Those who cannot remember the past are condemned to repeat it."

–George Santayana, The Life of Reason

Learn from Aviation and Computer Security



NATIONAL TRANSPORTATION SAFETY BOARD

Search this site... Search Site

HOME NEWS & EVENTS SAFETY ADVOCACY INVESTIGATIONS DISASTER ASSISTANCE LEGAL ABOUT PUBLICATIONS

Home > INVESTIGATIONS > Accident Reports > Aviation Accident Reports

SHARE

Aviation Accident Reports

The NTSB issues an accident report following the investigation. These reports are available online for reports issued since 1996, with older reports coming online soon. The reports listing is sortable by the event date, report date, city, and state. Click on any of those headings to sort the data.

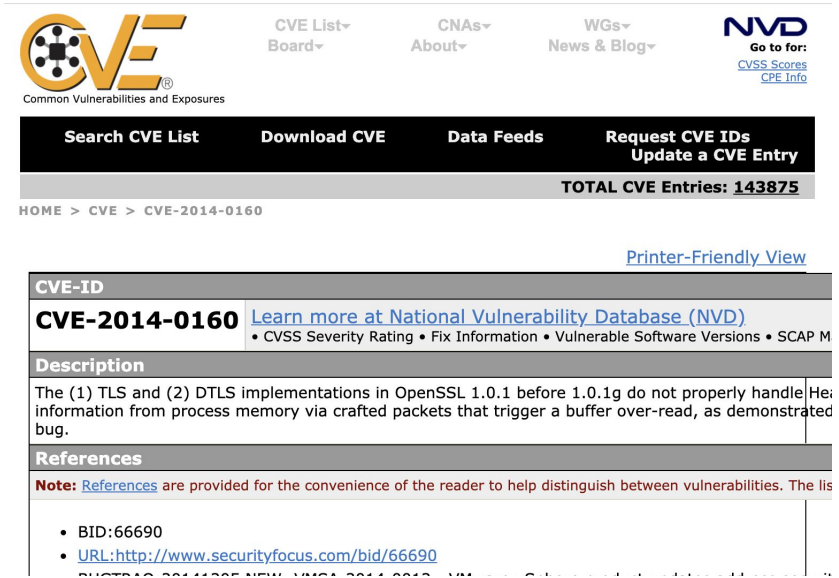
Showing 1 to 10 of 518 entries

First Previous 1 2 3 4 5 Next Last

Report Number	NTSB Title	Accident Date	Report Date	City	State	Country	NTISNumber	Report
AAR-20-02	Rapid Descent and Crash into Water Atlas Air Inc. Flight 3591 Boeing 767-375BCF, N1217A	2/23/2019	7/14/2020	Trinity Bay	TX	USA	PB2020-101004	PDF
ASR-20-04	Install Flight Data, Audio, and Image Recorder Systems on all Turbine-Powered Helicopters	7/1/2017	5/19/2020	Multiple	Multiple	USA		PDF
AAR-20-01	Helicopter Air Ambulance Collision with Terrain Survival Flight Inc. Bell 407 Helicopter, N191SF	1/29/2019	5/19/2020	Zaleski	OH	USA	PB2020-101001	PDF
ASR-20-02	Safety Recommendation Report: Revise Processes to Implement Safety Enhancements for Alaska Aviation Operations		2/13/2020		AK	USA		PDF
ASR-20-01	Reported Flight Control System Difficulty on Embraer EMB-175	11/6/2019	1/16/2020	Atlanta	GA			PDF
AAR-19-04	Inadvertent Activation of the Fuel Shutoff Lever and Subsequent Ditching Liberty Helicopters Inc., Operating a FlyNYON Doors-Off Flight Airbus Helicopters AS350 B2, N350LH	3/11/2018	12/10/2019	New York	NY	USA	PB2020-100100	PDF
AAR-19-03	Left Engine Failure and Subsequent Depressurization Southwest Airlines Flight 1380 Boeing 737-7H4, N772SW	4/17/2018	11/19/2019	Philadelphia	PA	USA	PB2019-101439	PDF
ASR-19-01	Safety Recommendation Report: Assumptions Used in the Safety Assessment Process and the Effects of Multiple Alerts and Indications on Pilot Performance	10/29/2018	9/19/2019	Multiple		Multiple		PDF
AAR-19-02	Departure From Controlled Flight Trans-Pacific Air Charter, LLC Learjet 35A, N452DA Teterboro, New Jersey May 15, 2017	5/15/2017	3/12/2019	Teterboro	NJ	USA	PB2019-100271	PDF
AAR-19-01	Runway Overrun During Rejected Takeoff Ameristar Air Cargo, Inc., dba Ameristar Charters, flight 9363 Boeing MD-83, N786TW Ypsilanti, Michigan, March 8, 2017	5/8/2017	2/14/2019	Ypsilanti	MI	USA	PB2019-100293	PDF

Showing 1 to 10 of 518 entries

First Previous 1 2 3 4 5 Next Last



CVE List Board CNAs About WGs News & Blog NVD Go to for: CVSS Scores CPE Info

Common Vulnerabilities and Exposures

Search CVE List Download CVE Data Feeds Request CVE IDs Update a CVE Entry

TOTAL CVE Entries: 143875

HOME > CVE > CVE-2014-0160

Printer-Friendly View

CVE-ID

CVE-2014-0160 Learn more at National Vulnerability Database (NVD)

• CVSS Severity Rating • Fix Information • Vulnerable Software Versions • SCAP M

Description

The (1) TLS and (2) DTLS implementations in OpenSSL 1.0.1 before 1.0.1g do not properly handle Hea information from process memory via crafted packets that trigger a buffer over-read, as demonstrated bug.

References

Note: References are provided for the convenience of the reader to help distinguish between vulnerabilities. The lis

- BID:66690
- URL:http://www.securityfocus.com/bid/66690

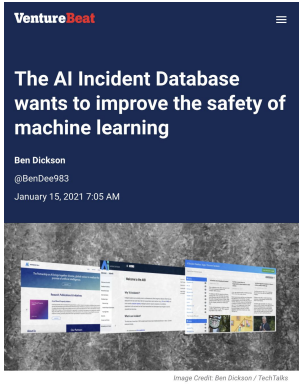
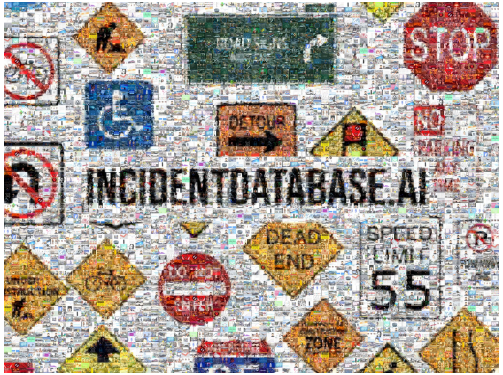


Image Credit: Ben Dickson / TechTalks



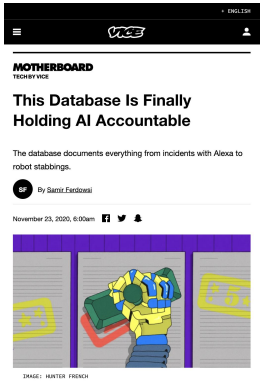
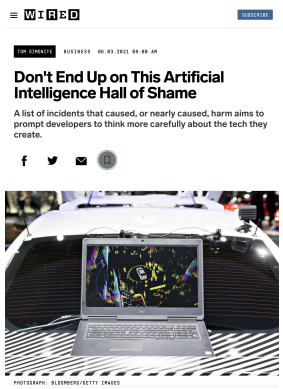


IMAGE: HUNTER FRENCH



PHOTOGRAPH: BLOOMBERG/GETTY IMAGES

1. Launched publicly in November
2. A project of the Partnership on AI
3. Current users: system architects, industrial product developers, public relations managers, researchers.

Demo: Incident Listing

Introducing the AI Incident Database

Basic idea: Help people discover how AI can go wrong

Demo: Discover Application

Introducing the AI Incident Database

Basic idea: Build community around AI Incident reporting and research

Demo: Submission and Leaderboard

Making a Community Through Citations

Demo: [Citation Pages]

https://incidentdatabase.ai/cite/4

https://incidentdatabase.ai/cite/92

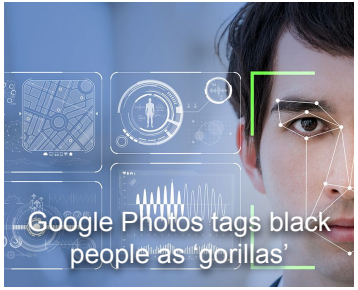
https://incidentdatabase.ai/cite/72

https://incidentdatabase.ai/cite/40

https://incidentdatabase.ai/cite/93

https://incidentdatabase.ai/cite/37

https://incidentdatabase.ai/cite/91



https://incidentdatabase.ai/cite/16

https://incidentdatabase.ai/cite/35

https://incidentdatabase.ai/cite/57

AIID Roadmap

1. (Done) Initial incident collection
2. (Done) Discover application development
3. (60 percent) Flexible taxonomy support
4. (Future) Incident monitoring
    a. Scrape entire internet on daily basis
    b. Support for all automatically-translatable languages

Taxonomy Support

Basic idea: AIID provides a foundation (incident reports) on which different groups can build

Demo: CSET Taxonomy

Open Source for Global Collaboration

Incident Infrastructure

PARTNERSHIP ON AI

Qualitative Contributions

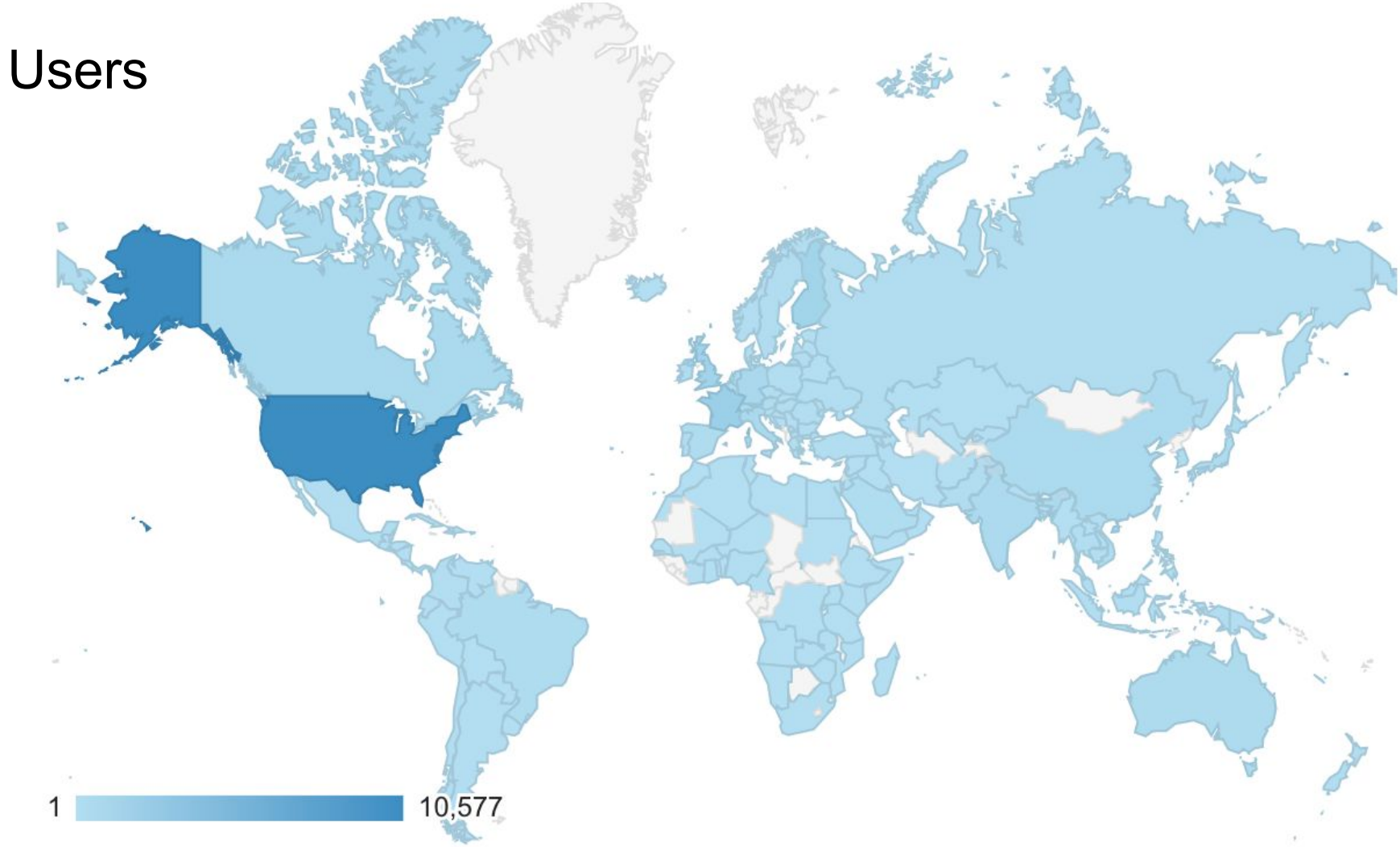
You?

Demo: Open Source

The "Bold Claims Slide"

- AI incident rates are accelerating (first evidence soon)
- Severe AI incidents are getting more common (first evidence soon)
- The vast majority of incidents are avoided with human-centered AI
- The AIID is critical to motivating the adoption of human-centered AI

Users



Summary

- Few people would fly today if not for a century of learning from aviation failures
- Submit incidents!
- Learn from the past!
- Use the AIID to recenter your intelligent systems on the humans

incidentdatabase.ai

PARTNERSHIP ON AI

Summary

- Few people would fly today if not for a century of learning from aviation failures
- Submit incidents!
- Learn from the past!
- Use the AIID to recenter your intelligent systems on the humans
- "Good" AI is more profitable than bad

Google text suggestion agrees!

incidentdatabase.ai

PARTNERSHIP ON AI

Thanks

Demo: Governance