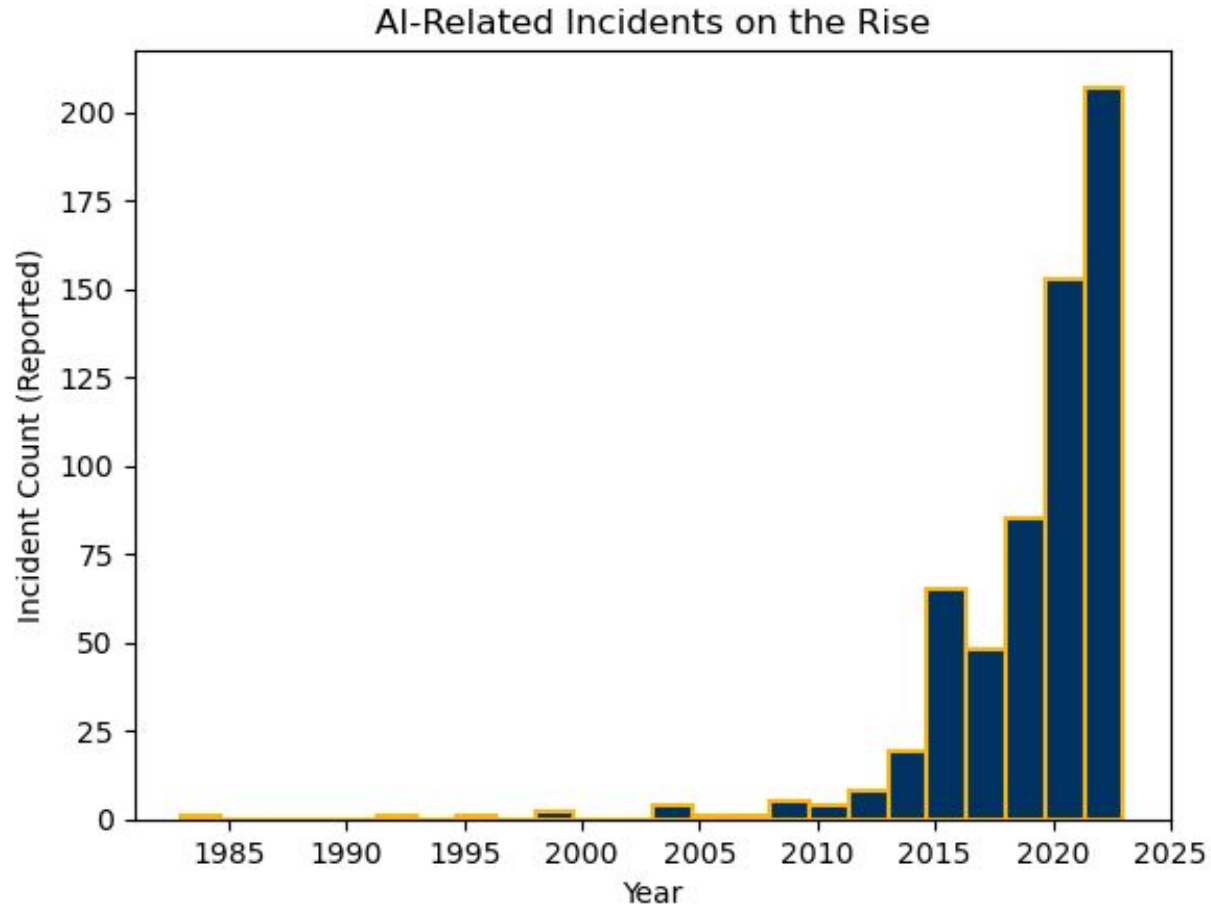


Towards Responsible AI: Rethinking Incident Management and Accountability in Artificial Intelligence



Jared Feldman
December 2023

Introduction

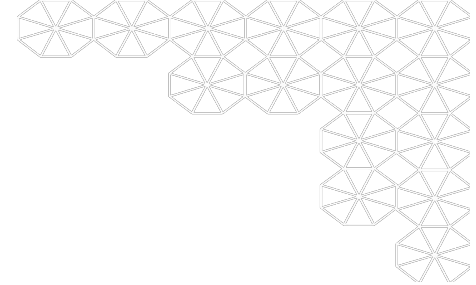


Artificial Intelligence Incident Database (AIID)

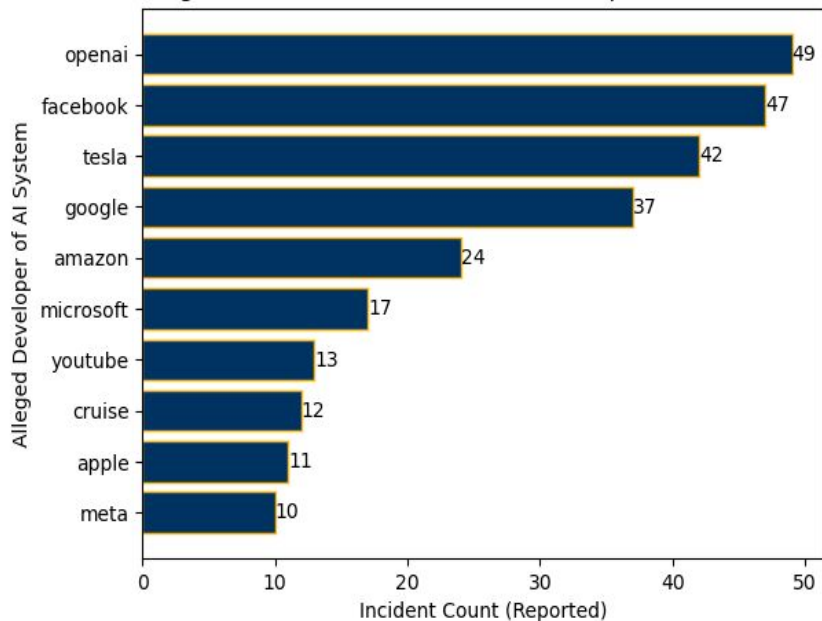


- **Incident:** an alleged harm or near harm event to people, property, or the environment where an AI system is implicated
- **Developer of AI System:** the organizations or individuals responsible for producing either the parts or the whole intelligent system implicated in the incident.
- **Harmed/Nearly Harmed Parties:** impacted classes (e.g., teachers, black people, women) or individuals.

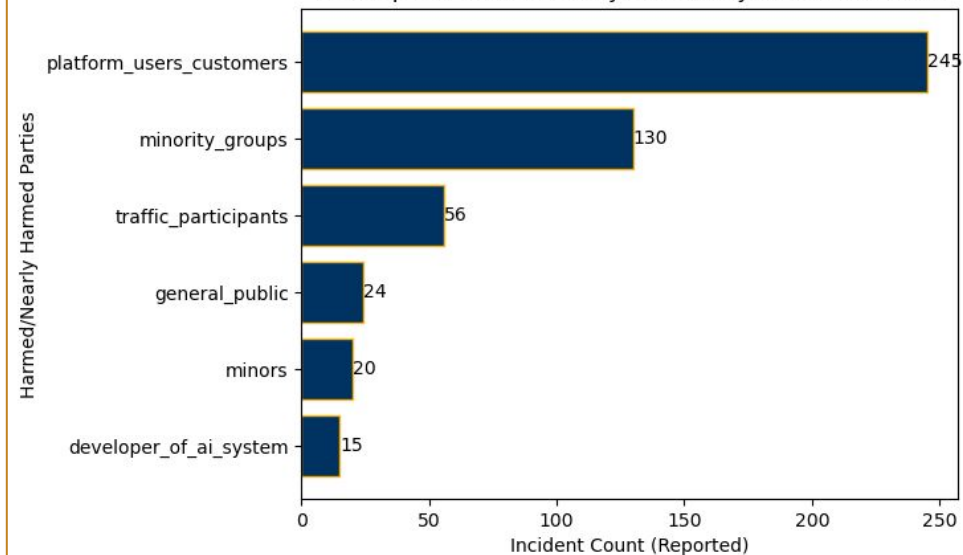
AIID: Trends and Concerns




10 Organizations Account for 43% of All Reported AIID Incidents



6 Groups Harmed or Nearly Harmed by 81% of Incidents



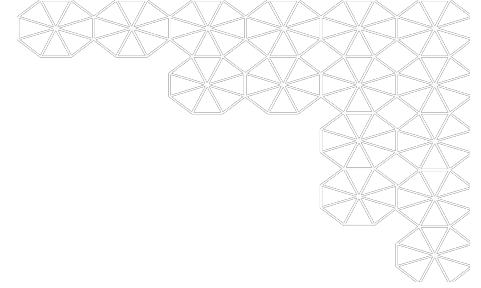
Accountability



*“The report reveals several unmistakable facts: the MAX crashes were not the result of a singular failure, technical mistake, or mismanaged event. They were the horrific culmination of a series of **faulty technical assumptions by Boeing’s engineers, a lack of transparency on the part of Boeing’s management, and grossly insufficient oversight by the FAA**—the pernicious result of regulatory capture on the part of the FAA with respect to its responsibilities to perform robust oversight of Boeing and to ensure the safety of the flying public. The facts laid out in this report document a disturbing pattern of technical miscalculations and **troubling management misjudgments made by Boeing**. It also illuminates numerous oversight lapses and **accountability gaps by the FAA** that played a significant role in the 737 MAX crashes.”*

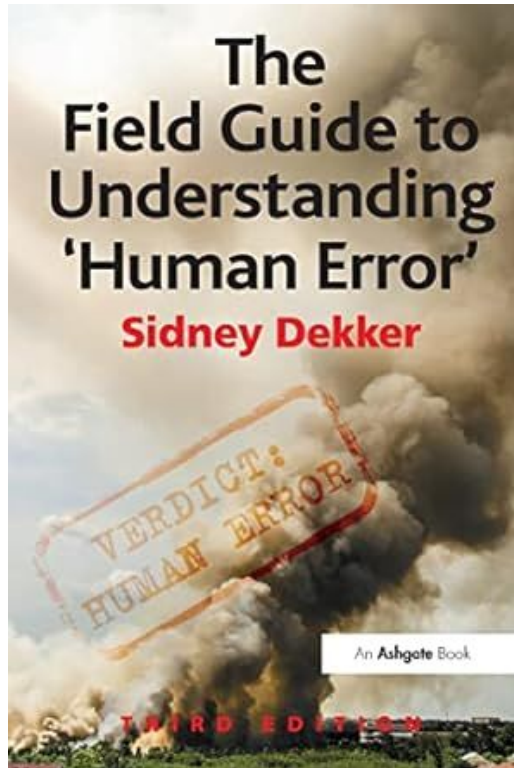
- Findings of the the House Committee on Transportation and Infrastructure on the 737 MAX incidents

Distributed Responsibility

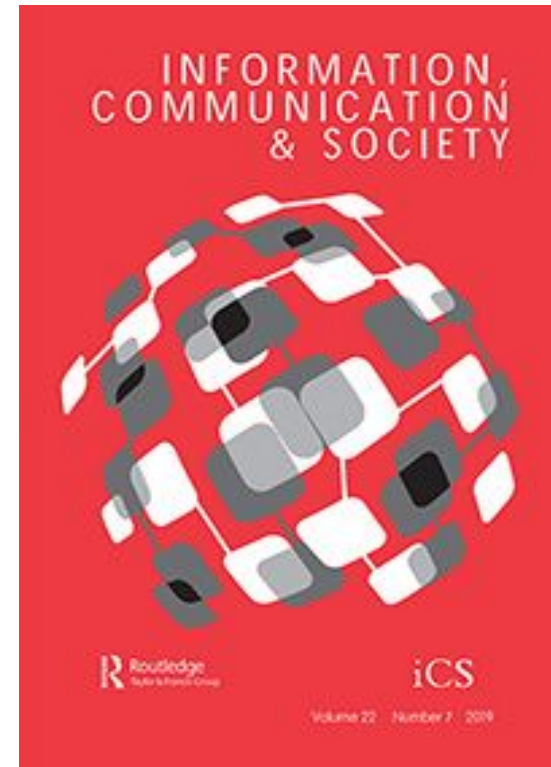


- Individual Contributors
- Ethics Owners
- Mid-Level Management/Leadership
- Senior Leadership/Executives
- The AI System
- Regulators
- Users

But... what about the idiots? Related Work on Accountability

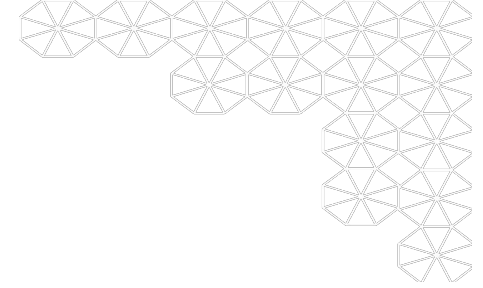


The Field Guide to Understanding “Human Error”
by Sidney Dekker



Where Fairness Fails
By Anna Lauren Hoffman

Proposed Path Forward



- Incident Management Processes
- Restorative Justice and Forward-looking Accountability
- Ethics and Privacy Owners
- Recalls

“Ethics can’t live on paper. They live inside people.”
Moss and Metcalf 2020

Thank you!

