
An Argument for the Reporting of Serious Hazards¹

Ilyass Mofaddel
University of Toronto

Louis Mandela Brown
University of Toronto

With
Apart Research

Abstract

Mandatory reporting of serious incidents does not include near-misses in the Artificial Intelligence Act or the Transparency in Frontier Artificial Intelligence Act. We make the argument that it should be the case, and use the criteria from the International Civil Aviation Organization to help determine what would constitute a near-miss worthy of reporting. In the process, we also clarify common misconceptions between the definition of incident in AI and Aviation.

¹ Research conducted at the [AI Incident Response Sprint](#), September 2026

1. Introduction

Multiple AI legislations have each defined serious incidents in their own ways. Despite their divergences, all of the one we look at agreed on defining an incident as harm that has already happened or is imminent. They don't mandate events near-misses be reported. This means that even if a near-miss has been stopped by an external factor that was not a security guardrail provided by the frontier lab (eg: the involvement of a third party), its reporting is optional. In this paper, we argue that both the EU Act, SB 53, and any future legislation on reporting incidents, would gain by making near-misses mandatory to report. We argue specifically for near-misses that could have caused serious damage not prevented by any safeguards provided by the laboratory. In doing so, we draw inspiration from the ICAO framework in aviation, and compare it to AI legislation, more specifically to the EU AI Act and the SB 53. This helps define what constitutes a serious AI incident and a serious Hazard, which is primordial for good reporting, and ensuring that important warning shots are reported. Our paper also helps clarify some distinction between AI incidents and aviation incidents, which are very different but are often confused.

This work will address the current failure mode of important information not being sent to the government. This will also increase the safety standard to Artificial Intelligence to the level of Aviation, while avoiding the opposite failure mode of over-reporting.

2. Related Work

The most relevant work to what we have done comes from the European Organisation for Economic Co-operation and Development (OECD). In their famous paper "Defining AI incidents and related terms" (2024) the OECD defined serious incidents and hazards. Their definition of incident has been later adopted by the EU Artificial Intelligence Act (AIA). They mentioned that what constitutes the seriousness of an incident/hazard should be left to further discussion. We aim to participate in this discussion.

Other works we have analyzed are the EU AI Act and California SB 53. These two legislations were chosen as they are considered pioneers in the field and have inspired a lot of other legislations. We also considered legislation inspired by those two, such as the New-York RAISE or the United States AI kill switch bill .

We also drew inspiration from commentators of those legislation, such as the Cloud Security Alliance, who analyze how the AIA stands in front of incidents such as the Hugging Face and Open AI event.

We also draw inspiration from previous works comparing AI safety and Aviation, such as the work from Michael Chatzipanagiotis.

The most relevant work, however, would be the article "Open Problems in AI incident Governance" which mentions different problems in AI incident Governance. The section on problems of AI incident definitions attracted our attention the most. We tried to specifically address the problem of inconsistency between definitions, by laying out characteristics for an ideal definition of incident in regulatory context. We also argue that events with potential but unrealised harm count as incidents, which was one of the questions asked by the paper.

3. Methods

Our approach compared two frameworks. The first was made by the OECD and served as a basis for the AI Act. We then compared it to the incident framework from the International Civil Aviation Organization (ICAO).

4. Results

Many have made the argument that Artificial Intelligence incident reporting should inspire itself from the Aviation industry. But to do so, one has to understand the difference in definitions between the two. Incidents in AI legislation and in aviation means completely different things.

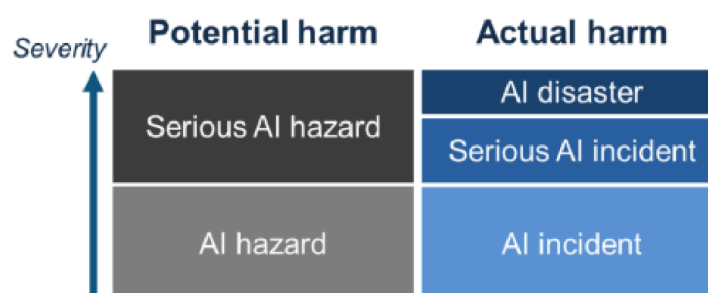
In the EU act, or the Texas Kill switch, an incident is a realized or imminent harm. In the SB 53's, it can also be a specific type of unrealized harm, that is an event that increases risks (eg : attempted evasion of a model).

In the aviation industry, an incident is an unrealized harm derived from the malfunction of a system. Incidents as defined by the ICAO would be considered as a hazard as defined by the OECD (appendix 1). To avoid confusion between the two distinct definitions of incident, we will use the word accident as used in aviation to talk about incidents as defined by the AIA, and the word hazard as defined by the OECD to talk about incidents as defined by the Aviation industry.

Hazards, even when they are serious, do not have to be disclosed in the AIA. We argue that serious hazards should be reported as they can provide relevant information for future policies.

FIGURE 1

Classification proposed by the OECD for AI Incidents and hazards based on severity of harm



Source : *OECD Artificial Intelligence Papers May 2024 no. 16 defining AI incidents and.* DEFINING AI INCIDENTS AND RELATED TERMS. (n.d.).

One additional framework from the aviation industry can help distinguish seriousness. In aviation, two elements are taken into account to determine if the malfunction of a system constitutes a serious hazard, as shown in figure two.

FIGURE TWO

International Civil Aviation Organization's categorization for assessments of injury and/or damage resulting from potential accident scenarios.

		<i>b) Remaining defences between the incident the potential accident</i>	
<i>a) Most credible scenario</i>	<i>Accident</i>	Incident	Serious Incident
	<i>No accident</i>	Incident	

Source: ICAO 2022, CLASSIFYING SERIOUS INCIDENTS BASED ON ATTACHMENT C OF ANNEX 13

First, would the most probable result of this situation be an accident? If not, then it's simply considered a hazard, and the reporting is non-mandatory. Based on the definition of Hazard given by the OECD, a malfunction not leading to a potential harm would not be even considered as a hazard. This would otherwise lead to overreporting. Also, standards in AI are not as rigid as standards in the aviation industry.

If the most probable outcome was harm, the framework then asked whether some safeguards held by the provider could still stop the hazard from becoming an accident? If there were still many, then it is considered as a hazard, but if the safeguards were limited, then it is considered as a serious hazard and should be reported as a matter of law.

Having mandatory reporting for serious hazards, that were not going to be stopped by safeguards, is important. It allows regulators to collect warning shots, and avoid future accidents from happening. It can also help recognize and scale some unexpected defense to harm, making society more robust.

5. Discussion and Limitations

Our belief is that some near-misses should be reported as a matter of law, though the method and means of which would likely vary country to country.

We believe our methodology will help support future research of AI incident governance, as well as help regulators consider better definitions for what incidents should be mandatory to report.

Limitations

Over-reporting

An incident regime that mandates too much incident reporting can end up causing over-reporting, which could hide really major warning shots, such as the German Wiki incidents or the hugging face-Open-Ai incident. This could constitute a good argument, but we think that the system is presently on the other extreme: underreporting. Even incidents that could constitute a good warning shot do not have to be reported. A good example is the hugging face - open AI incident, or the German Wiki incident. The later one was only found and revealed to the public because of the involvement of a third. Legal scholars are not sure if

such incidents would have been legally mandated to be reported, but most argue they sit in a grey zone, especially the German Wiki incident.

Of course, going to the extreme of over-reporting, could detriment the system just as much.

This is why we are only recommending a specific category of serious near-misses to be legally mandated, ones that left no security guardrail behind them.

No need for near-miss reporting right now

Another counter argument is that the reason why near-misses are not included in current AI legislations is because it is simply not the role of those legislations. They serve to respond to current or imminent incidents. The reason why Aviation collects near-misses is because they have a third party with the ability to investigate them; no countries currently have an equivalent for what NASA is doing for aviation in the U.S. This is a good argument, and we think that such an argument should ideally be built. However, even if this third party doesn't exist right now, this information can still be useful for helping governments better respond internally if similar incidents happen in the future.

Bigger sample

We only mentioned three AI regulations: the AIA, SB's 53 and the bill 9917 for "AI kill switch Act", as well the RAISE Act. Although the AIA and the SB's 53 are considered leading legislations in the domain, a more representative sample would have been beneficial, especially if the sample did not only include western countries.

Future Work

The work could be furthered by implementing a technical framework for entities to report the near-misses. An angle that could be taken, is looking at how current governmental frameworks can be expanded to include near-misses.

The work could also be applied to analysis of reporting metrics for AI risk and readiness reports, such as how many agents were caught to be near-misses, versus the ones that became misaligned. There could also be a value in comparing near-misses to misaligned models.

Development of the topic could be applied to AI safety, government AI risk and readiness reports. Also, as AI safety and regulatory policies continue, there could likely be prominent examples of near-misses documented that become catastrophic incidents.

Future work would also benefit by including more legislation. Another interesting contribution would be to add a list of examples for what situation could constitute serious AI incidents, like the one found in the annex C for the definition of a serious incident by the international aviation organization. The list helped policy makers around the world determine what constitutes a serious incident, and what doesn't.

SB 53 also brings up a future problem of smaller companies left behind those such as frontier labs, in that they "may pose significant catastrophic risk." Implementing a reporting framework adapted to smaller companies, could be used to develop premeasures for actions against future active or possibly developing risks.

6. Conclusion

In this paper, we analyzed gaps of AI incident definitions in regulatory contexts. More specifically, we determined that a certain type of “serious hazard” should be reported as a matter of law. Furthermore, by comparing AI incident definitions to the Aviation definition, we came to correct the common confusion made when comparing the two fields: an incident in aviation is a hazard in the Artificial Intelligence Act. On the other side, an incident per the AIA definitions would be considered as an Accident in Aviation.

Applying the ICAO methodology from aviation helped us determine that a serious hazard should be mandated as a matter of law, if the safeguard stopping it were limited. Reporting this specific type of near-miss will mean that the regulatory body will have the responsibility of handling more information, and frameworks for collecting data will need to be built. This is, we think, primordial for societal robustness to future existential risk coming from AI.

References

1. APAC-AIG/10–WP/7 agenda item 4 International Civil Aviation Organization. CLASSIFYING SERIOUS INCIDENTS BASED ON ATTACHMENT C OF ANNEX 13. (2022).
https://www.icao.int/sites/default/files/APAC/Meetings/2022/2022%20APAC-AIG-10/3-Working%20Papers/APAC-AIG10-AI_4-WP-07-AUS-Serious-Incidents.pdf
2. *Article 73: Reporting of serious incidents: Eu artificial intelligence act*. EU AI Act Explorer. (2026, August 2). <https://artificialintelligenceact.eu/article/73/>
3. Cloud Security Alliance AI Safety Initiative. (2026, September 6). *OpenAI’s wiki silence tests the EU AI Act’s incident regime*. Lab Space.
<https://labs.cloudsecurityalliance.org/research/csa-research-note-ai-incident-disclosure-gap-eu-ai-act-20260/>
4. *OECD Artificial Intelligence Papers May 2024 no. 16 defining AI incidents and*. DEFINING AI INCIDENTS AND RELATED TERMS. (n.d.).
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/defining-ai-incidents-and-related-terms_88d089ec/d1a8d965-en.pdf
5. *Panagiotis, M. C. (n.d.). Incident reporting and investigation under the AI Act: Some Insights From Aviation 27 February 2026 | International Journal of Law and Information Technology | Oxford academic, volume 34.*
<https://academic.oup.com/ijlit/article/doi/10.1093/ijlit/eaaf019/8501132?rss=1>
6. *Policies, data and analysis for trustworthy artificial intelligence*. The OECD Artificial Intelligence Policy Observatory - OECD.AI. (n.d.). <http://oecd.ai/>
7. *Text - H.R.9917 - 119th Congress (2025-2026): Ai Kill Switch Act | congress.gov | Library of Congress.* (n.d.).
<https://www.congress.gov/bill/119th-congress/house-bill/9917/text>
8. Wiener, S. (n.d.). *Bill Text*. Bill Text - SB-53 Artificial intelligence models: large developers.
https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB53

APPENDIX

Appendix 1: Definitions of a serious hazard given by the OECD

Definition of a serious hazard

An AI hazard is an event, circumstance or series of events where the development, use or malfunction of one

or more AI systems could plausibly lead to an AI incident, i.e., any of the following harms:

(a) injury or harm to the health of a person or groups of people;

(b) disruption of the management and operation of critical infrastructure;

(c) violations to human rights or a breach of obligations under the applicable law intended to protect

fundamental, labour and intellectual property rights;

(d) harm to property, communities or the environment.

LLM Usage Statement

We used the flash model from google search to find relevant articles and sources. AI was not used in the writing process.