

The Embodiment Gap: What Mandatory, Regulatory and Media-Based Incident Regimes Reveal About Robotic AI Failures, 2021–2026

Alfredo Merlet

alfredo.merlet@gmail.com

Universidad del Alba <https://orcid.org/0009-0007-8558-164X>

Research Article

Keywords: robotics incidents, incident reporting, automated driving systems, surgical robots, humanoid robots, risk regulation, AI governance

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11394165/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Robots driven by artificial intelligence now fail in public: robotaxis block ambulances, surgical instruments break inside patients and humanoid robots lose control at exhibitions. What society learns about these failures depends on which incident regime records them. This article compares three regimes for embodied AI between 2021 and 2026: the mandatory crash-reporting order of the US National Highway Traffic Safety Administration (8,135 incidents involving automated driving systems or Level 2 driver assistance), the Food and Drug Administration's adverse-event register for computer-assisted surgical systems (127,627 reports since 2015) and the media-derived AI Incident Database (1,734 incidents). Fifty-four road-vehicle incidents in the media-derived database were hand-coded and linked to mandatory reports. Twenty of the 21 US crashes in the media-derived database were also reported to the regulator, but only 21 of 8,135 mandatorily reported incidents (0.26%) reached the media-derived database. Contrary to the pre-specified expectation, media visibility rose with severity: 3.1% of fatal or serious incidents against 0.37% of the rest (odds ratio 8.7; Fisher exact $p < 0.001$), a gradient that keeps its direction across nine robustness checks. Almost half of the media-recorded US events (19 of 40) were operational failures without a crash, which a crash-triggered order cannot capture by design. In the surgical register, a fourfold rise in malfunction reports in 2025 is concentrated in two wear-related problem codes filed by the manufacturer, while injury reports barely changed. Humanoid robots, the most visible failures of 2026, fall under no reporting regime. Each regime sees a different class of failure, and none sees the robots now entering public space.

1. Introduction

Artificial intelligence has acquired a body. In 2026, driverless robotaxis carry passengers in a growing number of US cities, computer-assisted surgical systems are routine in urology and gynaecology, and humanoid robots have moved from laboratories to trade fairs, amusement parks and city streets. Unitree, one of the largest Chinese manufacturers, announced plans to ship up to 20,000 humanoid robots in 2026 (eWeek, 2026). The same year produced a sequence of public mishaps: a remotely operated humanoid that frightened an elderly woman in Macau and was escorted away by police (Yahoo Tech, 2026), a humanoid that embraced a dancer during a university performance in Xi'an (Global Times, 2026), a robot that reportedly kicked a boy during a demonstration in Urumqi (AI Incident Database [AIID], 2026), a humanoid that caught fire after crashing into a barrier at the World Humanoid Robot Games in Beijing (KTLA, 2026), and a humanoid seized by German police after allegedly touching a 16-year-old girl in Cologne (CBS News, 2026).

These events reach public knowledge through different channels. Some are reported under a legal obligation, others are filed by manufacturers to a health regulator, and others exist only because a journalist or a bystander published a video. The arrangement that decides which failures are recorded, by whom and under which trigger is what this article calls an incident regime. In the vocabulary of risk regulation, it is the information-gathering component of a regime, alongside standard setting and behaviour modification (Hood et al., 2001). Its design matters because policymakers, insurers, researchers and the public reason about the safety of embodied AI from whatever the regimes happen to record.

Two bodies of research study robotic incidents, but they rarely meet. Safety research on automated vehicles and surgical robots analyses one mandatory or regulatory register at a time, such as crash reports filed by robotaxi operators (Kusano et al., 2024, 2025; Wang et al., 2026) or adverse-event reports for the da Vinci system (Alemzadeh et al., 2016; Hsieh & Huang, 2026). Research on AI incidents, in contrast, pools voluntary, media-derived collections such as the AI Incident Database (Wei & Zhou, 2023; May et al., 2024; Tuovinen, 2025; Raghupathi et al., 2026). The editors of that database acknowledge that uncertainty about cause, severity and scope is unavoidable in AI incident reporting (Paeth et al., 2025). What has not been measured is how much a media-derived collection misses relative to a mandatory register for the same technology, and whether what it misses is random or systematic. A Scopus search run

on 15 October 2026 returned four records at the intersection of robotic incidents, reporting systems and comparative governance, none of which compares regimes (Appendix A).

This article asks what each incident regime sees of robotic AI failures, and how the regimes differ in the class and severity of the failures they record. It answers with three public sources that cover the period 2021–2026: the Standing General Order of the US National Highway Traffic Safety Administration (NHTSA), which obliges named companies to report crashes involving automated driving systems (ADS) and Level 2 driver-assistance systems; the Manufacturer and User Facility Device Experience (MAUDE) register of the US Food and Drug Administration (FDA), queried for computer-assisted surgical systems; and the AI Incident Database (AIID), which catalogues incidents documented in published reports.

The article makes three contributions. First, it links records across a media-derived database and a mandatory register for the same technology, which yields the first estimate of media visibility conditional on harm severity for robotic AI. Second, it decomposes a sudden change in the volume of a regulatory register to show how much of it reflects the behaviour of the reporter rather than the evolution of harm. Third, it documents that humanoid robots, the newest and most visible form of embodied AI, are covered by none of the three regimes. The distance between where embodied AI fails and where those failures are recorded is what the title calls the embodiment gap.

Section 2 reviews the literature on incident regimes and on each of the three channels and states the gap. Section 3 sets out the hypotheses. Section 4 describes the data, the coding and linkage procedure and the statistical methods. Section 5 reports the results, Section 6 discusses them and Section 7 concludes.

2. Incident regimes and the literature on robotic failures

2.1 Incident registers as information-gathering systems

Hood et al. (2001) describe risk regulation regimes as systems with three control components: gathering information, setting standards and modifying behaviour. Information gathering is not a neutral mirror of risk. What a register contains depends on its trigger (which events must be reported), its reporters (who knows about the event and has an obligation or an incentive to report it) and its filters (who decides whether a report counts). Safety science reached a similar conclusion from the side of organisations. Reason (1997) argued that reporting cultures determine which near misses become visible, Vaughan (1996) showed how deviations become normal when they are not recorded as anomalies, and Leveson (2011) located accidents in the control structure of a socio-technical system rather than in the failure of a single component. Perrow (1984) had already warned that tightly coupled, complex systems produce accidents that no single register anticipates. Power (2007) adds that risk management produces auditable representations of risk, which then become the object of management in their own right.

For emerging technologies, the information-gathering component carries a particular burden. Collingridge (1980) described a dilemma: early in the life of a technology, its harms are hard to know, and by the time they are known, the technology is entrenched and hard to change. Incident registers are the main instrument for breaking that dilemma, because they accumulate evidence about harm while the technology is still being deployed. Their coverage therefore shapes how early regulators learn.

2.2 The mandatory channel: crash reporting for automated vehicles

Since June 2021, NHTSA's Standing General Order 2021-01 has required the manufacturers and operators it names to report crashes in which an ADS or a Level 2 system was engaged within 30 seconds of the crash. The criteria differ by system: ADS crashes are reportable if they cause certain property damage or injury, while Level 2 crashes are reportable

if they involve a fatality, a struck vulnerable road user, an air-bag deployment or a hospital transport (National Highway Traffic Safety Administration [NHTSA], 2026). A third amendment, in force since 16 June 2025, streamlined the order: the most severe crashes are due within five days, less severe ones monthly, and multiple entities no longer report the same crash unless their information differs (NHTSA, 2025). The agency itself warns that the data are not normalised by fleet size or distance travelled and that companies with richer telematics will report more crashes simply because they know about more of them (NHTSA, 2026).

Research on automated-vehicle safety uses these and similar sources. Favarò et al. (2017) analysed collision reports filed with the California Department of Motor Vehicles; Wang et al. (2026) applied latent class analysis and interpretable machine learning to Level 4 crashes in California; Kusano et al. (2024, 2025, 2026) compared Waymo's rider-only crash rates with human benchmarks; Mimi et al. (2026) classified crash severity by automation level using Texas crash narratives; and Patel et al. (2026), Astle and Haus (2026) and Shah (2026) examined disengagements, roadway context and injury outcomes. Kalra and Paddock (2016) showed why such registers cannot settle questions of relative safety on their own: demonstrating a reduction in fatality rates requires exposure measured in hundreds of millions of miles. All of these studies treat a single register as the population of interest.

2.3 The regulatory channel: adverse events in robotic surgery

Under the US medical device reporting rule, manufacturers and user facilities must report deaths, serious injuries and certain malfunctions associated with devices, and anyone else may report voluntarily (21 C.F.R. pt. 803). The resulting MAUDE register has been the main source for studies of robotic surgery safety. Andonian et al. (2008) and Lucas et al. (2012) reviewed device failures and their clinical consequences; Friedman et al. (2013) focused on instrument failures; Alemzadeh et al. (2016) analysed fourteen years of reports and counted deaths, injuries and malfunctions, and earlier proposed a systems-theoretic safety assessment of telesurgical robots (Alemzadeh et al., 2015); Gupta et al. (2017) developed a classification scheme for da Vinci events; Nik-Ahd et al. (2019), Tapper et al. (2019) and Correa et al. (2022) examined trends, causes and specialty differences; and recent studies cover a decade of reports (Hsieh & Huang, 2026) or seventeen years of prostatectomy events (Hammadeh et al., 2026). This literature routinely notes that passive surveillance under-reports. A systematic review of spontaneous reporting of adverse drug reactions found a median under-reporting rate of 94% across studies (Hazell & Shakir, 2006). The opposite problem, a register whose volume grows because of how reporters behave rather than because harm grows, has received less attention.

2.4 The voluntary channel: media-derived AI incident databases

The AI Incident Database was created to give intelligent systems the collective memory of failures that mature industries such as aviation already have, and it defines an AI incident as a situation in which AI systems caused, or very nearly caused, real-world harm (McGregor, 2021). Its editors index incidents documented in published reports and attach independent taxonomies; they report that ambiguity about cause, multiplicity and severity is a structural feature of the task (Paeth et al., 2025). Studies built on such collections find that service robots and autonomous driving are among the most frequent application areas (Wei & Zhou, 2023), that autonomous driving and process automation dominate safety and security incidents (May et al., 2024), that robots form one of five categories of fatal or near-fatal AI incidents (Tuovinen, 2025), and that topic models of several thousand records reveal stable thematic structures (Raghupathi et al., 2026).

A media-derived collection inherits the selection logic of the news. Research on news values shows that negativity, unexpectedness, magnitude and relevance to elite actors raise the probability that an event is reported (Galtung & Ruge, 1965; Harcup & O'Neill, 2017). The social amplification of risk framework adds that the attention an event receives depends on how it is processed by journalists, institutions and audiences, not only on its physical consequences (Kasperson et al., 1988). Neither literature predicts unambiguously whether a media-derived incident database will over-

or under-represent the most harmful events: magnitude favours severe events, while unexpectedness favours novel but harmless ones.

2.5 Robots without a register and the research gap

Humanoid and service robots operating in public space are not covered by any dedicated incident register comparable to the crash-reporting order or the device register. In the European Union, the Machinery Regulation sets conformity requirements for machinery placed on the market (Regulation (EU) 2023/1230), and the AI Act requires providers of high-risk AI systems to report serious incidents to market surveillance authorities (Regulation (EU) 2024/1689, Article 73). Whether a given humanoid falls within either regime depends on its classification and use.

The gap this article addresses therefore has four parts. The safety literature on automated vehicles and surgical robots analyses registers one at a time; the AI-incident literature pools voluntary collections without linking them to mandatory ones; no study has estimated how media visibility varies with harm severity for the same technology; and no study has decomposed reporter-driven changes in the volume of a robotic register or mapped the humanoid domain against existing regimes. The Scopus funnel reported in Appendix A confirms that the full intersection returns four records and that relaxing it returns single-register studies only.

3. Hypotheses

The hypotheses were fixed before the linkage was performed; H3 was formulated after the regulatory series was first inspected and is labelled exploratory.

H1 (class of failure). The regimes differ in the class of failure they record: media-derived records include operational failures without physical contact, which a crash-triggered mandatory register excludes by design.

H2 (severity). Among incidents recorded in the mandatory register, the probability of appearing in the media-derived database is lower for fatal or serious incidents than for less severe ones. The expectation rested on the unexpectedness criterion of news selection: low-harm but novel robotaxi behaviour could attract more coverage than conventional-looking collisions with serious consequences.

H3 (exploratory; reporter behaviour). In the regulatory register, large changes in the volume of reports reflect changes in reporter behaviour more than changes in harm.

4. Data and methods

4.1 Sources

Table 1 summarises the three regimes and the domain without a regime. The NHTSA and openFDA files were downloaded on 15 October 2026; the AIID file is the weekly export of 12 October 2026. All raw files are included in the replication package.

Table 1
Incident regimes for embodied AI compared in this study

Regime	Domain and source	Obligation and trigger	Unit and coverage	Records
Mandatory, telemetry-linked	Road vehicles with ADS or Level 2 systems, US; NHTSA Standing General Order 2021-01 (NHTSA, 2025, 2026)	Named manufacturers and operators; crash with the system engaged within 30 s; severity thresholds differ by system; third amendment in force from 16 June 2025	Crash incident; July 2021–September 2026	8,135 incidents (ADS 3,059; Level 2 5,076)
Regulatory, adverse-event	Computer-assisted surgical systems, product code NAY; FDA MAUDE via openFDA (U.S. Food and Drug Administration, 2026)	Manufacturers and user facilities report deaths, serious injuries and malfunctions; others may report voluntarily (21 C.F.R. pt. 803)	Device report; 2015–September 2026	127,627 reports
Voluntary, media-derived	All AI systems; AI Incident Database export of 12 October 2026 (AI Incident Database, 2026)	Editorial inclusion of incidents documented in published reports in which AI caused or nearly caused harm (McGregor, 2021)	Incident; all years	1,734 incidents; road vehicles 90, surgical robots 1, humanoid or legged robots 5
None identified	Humanoid robots in public and commercial settings	No dedicated incident register	–	Five 2026 cases compiled from news reports (Table 6)
Note: ADS, automated driving system (SAE Levels 3–5); Level 2, driver-assistance system providing combined steering and speed control. Source: author's compilation from the cited sources.				

4.2 Mandatory register

The NHTSA files contain one row per report version. The latest version of each report was retained, and reports were grouped into incidents with the agency's Same Incident ID, which links reports that describe the same crash. The ADS and Level 2 files of both the archive (second amended order, to 15 June 2025) and the current release (third amended order, from 16 June 2025) were combined; the file of reports whose automation type was not stated was excluded. Incidents were kept up to the incident month of September 2026, which yields 8,135 incidents: 3,059 involving an ADS and 5,076 involving a Level 2 system. Twenty-two incidents dated before July 2021 were reported retrospectively and are retained. The two versions of the order use different injury scales; they were harmonised into six levels (fatality, serious, moderate, minor, no injury and unknown) by merging the with and without hospitalisation variants of the current scale. Severity is unknown for 1.8% of ADS incidents and for 86.1% of Level 2 incidents, and 90.2% of Level 2 narratives are redacted as confidential business information. Tesla accounts for 87.4% of Level 2 incidents and Waymo for 2,318 of the 3,059 ADS incidents.

4.3 Media-derived database and coding

The AIID Excel export of 12 October 2026 contains 1,734 incidents, the latest dated 3 October 2026. Robotic domains were identified with regular expressions over the title, description, deployer and developer of each incident. For road vehicles, every incident dated from July 2021, when the mandatory order began, was read in full; 54 concern ADS or Level 2 vehicles. Each was coded on four variables: country; system (ADS or Level 2); event class; and, for US crashes, the matching mandatory report. The event classes are crash (physical contact between the subject vehicle and a road user, vehicle, animal or object), operational failure (hazardous or harmful behaviour without contact, such as obstructing emergency vehicles, becoming immobilised in traffic or entering a prohibited area), aggregate (several events without a

single dated event, such as an investigation) and no event (a perception error without consequence). The coding table, with a justification for every decision, is part of the replication package.

4.4 Linkage

For each US crash, the mandatory register was searched for reports by the same reporting entity in the same incident month, then narrowed by city or state, crash partner and narrative keywords. A match is exact when a single candidate agrees on every available field, and probable when a single candidate agrees on entity, month and city but the narrative is redacted. Each mandatory incident then receives an indicator equal to one if it is linked to a media-derived record. The indicator is the basis for the visibility estimates.

4.5 Regulatory register

The openFDA device-event interface (U.S. Food and Drug Administration, 2026) was queried for product code NAY, computer-controlled surgical instruments; the five most frequent brand names under the code (EndoWrist, da Vinci Xi, da Vinci 5 and variants) confirm that it covers the da Vinci family. Annual counts by date of receipt were retrieved for event type, report source, type of report, summary-reporting exemption number and device problem code. Problem codes are not mutually exclusive, and the interface returns the 100 most frequent codes per year, so small counts in early years may be incomplete. The register was last updated on 7 October 2026; the 2026 figures cover January to September.

4.6 Statistical analysis

Media visibility for severity level s is the share of mandatorily reported incidents that are also in the media-derived database: $v(s) = m(s) / n(s)$, where $n(s)$ is the number of incidents at level s and $m(s)$ the number linked. Intervals are Wilson score intervals (Wilson, 1927). The main test compares fatal or serious incidents with all other incidents of known severity using Fisher's exact test. Nine robustness specifications follow: ADS only (R1) and Level 2 only (R2); unknown severity treated as non-severe (R3); probable matches removed (R4); the match whose reporter disputes that the system was engaged removed (R5); the second (R6) and third (R7) amended orders separately; a Mantel–Haenszel odds ratio pooled across systems (R8; Mantel & Haenszel, 1959); and a Cochran–Armitage test for trend across the five ordered levels of known severity (R9; Armitage, 1955). With 18 linked incidents of known severity, the estimates are read as contrasts with wide uncertainty, not as population rates. Because none of the regimes records exposure, no statement is made about the relative safety of any system.

4.7 Humanoid cases

Humanoid incidents of 2026 were compiled from reports by named news organisations and from the AIID itself, and each was checked against the AIID snapshot. The compilation is purposive and illustrative; it is not used in any statistical test.

4.8 Use of generative AI and reproducibility

Claude Opus 5.5 (Anthropic) was used under the author's direction to write the analysis code, process the data, search the literature, produce the figures, draft the text and make a first pass of the incident coding and record linkage. The author reviewed every coding and linkage decision, checked every reference against its primary source and is responsible for the content. All tables and figures are produced by a single script from the raw files (Data availability).

5. Results

5.1 What each regime records

The three regimes differ in volume by orders of magnitude. The mandatory order recorded 8,135 road-vehicle incidents, rising steadily through both versions of the order (Fig. 1). The regulatory register received 127,627 reports for computer-assisted surgical systems between 2015 and September 2026, of which 117,170 were malfunctions, 9,873 injuries and 584 deaths. The media-derived database holds 90 road-vehicle incidents, one surgical-robot incident (a 2015 collection of malfunctions) and five incidents involving humanoid or legged robots across all years; for 2026 alone it records twelve road-vehicle incidents, no surgical incident and one humanoid incident.

5.2 Classes of failure (H1)

Table 2 reports the coding of the 54 road-vehicle incidents in the media-derived database. Of the 40 single US events, 21 are crashes and 19 are operational failures without a crash, a share of 47.5% (95% Wilson interval 32.9–62.5%). The operational failures include a robotaxi that drove off during a police stop, vehicles that blocked a fire truck and an ambulance, fleets immobilised by lost connectivity or a city-wide power outage, a person trapped undetected in a vehicle's trunk, a robotaxi that remained immobile while vandals attacked it with a passenger inside, and an unoccupied vehicle swept into a flooded waterway.

Table 2
Coding of road-vehicle incidents in the AI Incident Database, July 2021–October 2026

Event class	US	Outside US or unknown	Total	US events linked to a mandatory report
Crash	21	4	25	20 (18 exact, 2 probable)
Operational failure without crash	19	2	21	0
Aggregate (several events)	6	0	6	–
No event (perception error only)	0	2	2	–
Total	46	8	54	20
Note: Outside US includes China (3), Taiwan (1) and Germany (1); two video-documented perception errors have no stated location. Source: author's coding of AI Incident Database (2026); linkage to NHTSA (2026).				

Twenty of the 21 US crashes were found in the mandatory register: 18 exact and two probable matches, both involving Level 2 vehicles with redacted narratives. The unmatched crash is a 2023 collision on the Pennsylvania Turnpike. In one matched case the reporting company states in its own report that the automated system was not engaged within the reporting window. None of the 19 operational failures has a counterpart in the register, as expected from its crash trigger; two of them involved physical damage to the robotaxi itself (the flooded roadway and a firework that set an unoccupied vehicle on fire) without a matching report. H1 is supported.

5.3 Severity and media visibility (H2)

Only 21 of the 8,135 mandatorily reported incidents (0.26%) appear in the media-derived database. Visibility rises monotonically with severity (Table 3 and Fig. 2): 3.8% of fatal incidents, 2.4% of serious, 1.1% of moderate, 0.7% of minor and 0.30% of incidents without injury, against 0.07% of incidents of unknown severity. Fatal or serious incidents were linked in 5 of 161 cases (3.1%, 95% interval 1.3–7.1%), and other incidents of known severity in 13 of 3,550 cases (0.37%, 0.21–0.63%). The odds ratio is 8.72 (Fisher exact $p = 0.0008$).

Table 3
Media visibility of mandatorily reported incidents by severity, July 2021–September 2026

Highest injury severity	ADS: linked / n	Level 2: linked / n	All: linked / n	Visibility (%)	95% CI (Wilson)
Fatality	0 / 3	3 / 75	3 / 78	3.85	1.32–10.71
Serious	1 / 10	1 / 73	2 / 83	2.41	0.66–8.37
Moderate	1 / 36	0 / 56	1 / 92	1.09	0.19–5.90
Minor	3 / 248	0 / 166	3 / 414	0.72	0.25–2.11
No injury	8 / 2,706	1 / 338	9 / 3,044	0.30	0.16–0.56
Unknown	1 / 56	2 / 4,368	3 / 4,424	0.07	0.02–0.20
Total	14 / 3,059	7 / 5,076	21 / 8,135	0.26	0.17–0.39
Note: linked means that the incident in the NHTSA register is matched to a record in the AI Incident Database. Severity levels harmonise the scales of the second and third amended orders. Source: author's calculations from NHTSA (2026) and AI Incident Database (2026).					

The gradient holds across the robustness specifications (Table 4). It is significant for Level 2 vehicles alone (odds ratio 15.5, $p = 0.008$) and borderline for ADS alone (20.7, $p = 0.055$), where only one severe incident is linked. The Mantel–Haenszel odds ratio pooled across systems is 16.5 (95% interval 2.7–100.4), and the trend test across the five ordered levels gives $z = 5.09$ ($p < 0.001$). Removing probable matches leaves the estimate unchanged, because both have unknown severity. Under the third amended order alone, only four linked incidents remain and the contrast is not significant ($p = 0.13$). H2, which predicted lower visibility for severe incidents, is not supported: the association runs in the opposite direction.

Table 4
Severe versus non-severe visibility: main estimate and robustness

Specification	Severe: linked / n	Non-severe: linked / n	Odds ratio	p-value
Main: ADS+ADAS, known severity	5 / 161	13 / 3,550	8.72	< 0.001
R1: ADS only	1 / 13	12 / 2,990	20.68	0.055
R2: Level 2 only	4 / 148	1 / 560	15.53	0.008
R3: unknown severity coded as non-severe	5 / 161	16 / 7,974	15.94	< 0.001
R4: probable matches removed	5 / 161	13 / 3,550	8.72	< 0.001
R5: disputed-engagement match removed	5 / 161	12 / 3,550	9.45	< 0.001
R6: second amended order only (to Jun-2025)	4 / 102	10 / 1,854	7.53	0.005
R7: third amended order only (from Jun-2025)	1 / 59	3 / 1,696	9.73	0.128
R8: Mantel-Haenszel, stratified by system	–	–	16.51	< 0.001
R9: Cochran-Armitage trend, five ordered levels	–	–	–	< 0.001
Note: severe = fatal or serious. p-values from Fisher's exact test except R8 (Mantel–Haenszel test of a common odds ratio of one; pooled odds ratio 16.51, 95% CI 2.71–100.41) and R9 (Cochran–Armitage test for trend, $z = 5.09$). Source: author's calculations.				

The fatal incidents that do not reach the media-derived database are informative. ADS operators reported three fatal crashes, none of which is in the database. In all three, the operator's narrative describes the robotaxi as stopped or slowing when another road user initiated the collision. The database's own definition requires that an AI system caused or nearly caused the harm (McGregor, 2021), so editorial attribution acts as a second filter after media attention.

5.4 Reporter behaviour in the regulatory register (H3, exploratory)

Malfunction reports for computer-assisted surgical systems rose from 11,016 in 2024 to 47,369 in 2025, an increase of 36,353 (Table 5 and Fig. 3). Injury reports moved from 1,778 to 1,833 and death reports from 69 to 98. The increase came from manufacturer reports, which rose from 94.7% to 99.1% of the total, while reports from user facilities fell from 449 to 311 and voluntary reports from 231 to 146. The summary-reporting exemption field is empty in every year. Two problem codes account for most of the change: Material Split, Cut or Torn rose from 694 to 26,867 reports and Material Frayed from 54 to 5,039. Together they equal 85.7% of the excess, an upper bound because codes can co-occur. Follow-up reports also rose, from 6,285 to 30,914. The pattern is consistent with H3: the manufacturer began reporting a specific instrument wear mode at scale, while reported harm changed little. Because the decomposition uses problem codes rather than narratives, it identifies where the change is concentrated, not why the manufacturer changed its reporting.

Table 5
Reports for computer-assisted surgical systems (product code NAY) in MAUDE, by year of receipt

Year	Malfunction	Injury	Death	Manufacturer share (%)	User facility	Voluntary	Split, cut or torn	Frayed	Unintended motion
2015	1,537	146	34	89.9	108	66	0	67	11
2016	890	170	17	84.7	109	56	0	16	0
2017	844	177	14	82.4	80	105	0	4	3
2018	893	156	13	80.8	86	120	0	3	1
2019	959	176	17	85.2	84	87	7	4	2
2020	1,175	344	35	87.3	113	85	49	4	4
2021	1,420	532	43	85.4	162	129	302	21	41
2022	4,346	1,275	100	95.8	114	125	313	76	342
2023	9,137	1,694	63	97.2	186	119	518	106	765
2024	11,016	1,778	69	94.7	449	231	694	54	800
2025	47,369	1,833	98	99.1	311	146	26,867	5,039	1,924
2026*	37,584	1,592	81	99.2	195	115	16,550	4,375	1,730

Note: *January–September 2026. Problem codes can co-occur within a report and are drawn from the 100 most frequent codes per year. Unintended motion sums all codes beginning with 'Unintended'. The summary-reporting exemption field is empty in all years. Source: author's calculations from U.S. Food and Drug Administration (2026).

Problem codes that describe unintended motion of the system (Unintended Movement, Unintended System Motion, Unintended Collision and related codes) offer a bridge to the loss-of-control events seen in humanoid robots. They appeared in at most 11 reports a year between 2015 and 2020, then in 342 reports in 2022, 800 in 2024, 1,924 in 2025 and 1,730 between January and September 2026.

5.5 Humanoid robots: a domain without a regime

Table 6 lists five humanoid incidents of 2026 reported by named outlets. Only one, the Urumqi demonstration, is in the media-derived database. The Macau and Xi'an incidents, which occurred six to seven months before the snapshot, are absent; the Beijing fire occurred seven weeks before it, and the Cologne case was reported after it, so their absence may reflect indexing delay. None of the five falls under a mandatory or regulatory reporting obligation comparable to those analysed above.

Table 6
Humanoid robot incidents reported in 2026 and their coverage

Date (2026)	Place	Event as reported	Source	In AIID snapshot
12 March	Macau, China	A remotely operated humanoid used in a promotion stopped behind a 70-year-old woman and could not move around her; police escorted it away and returned it to its owner with a warning	Yahoo Tech (2026)	No
April	Xi'an, China	A humanoid embraced a student during a choreographed university performance; the company attributed it to signal interference	Global Times (2026)	No
1 June	Urumqi, China	A humanoid identified as a likely Unitree G1 reportedly kicked a boy during a public demonstration	AI Incident Database (2026), incident 1536	Yes
26 August	Beijing, China	A humanoid caught fire after hitting the barrier at the end of the 100 m final of the World Humanoid Robot Games	KTLA (2026)	No (seven weeks before snapshot)
October	Cologne, Germany	Police seized a humanoid after it allegedly touched a 16-year-old girl; legal experts noted that harassment law requires direct contact between persons	CBS News (2026)	Not assessable (reported after snapshot)
Note: purposive compilation for illustration; not used in statistical tests. Snapshot: AI Incident Database export of 12 October 2026. Source: author's compilation from the cited sources.				

6. Discussion

6.1 Three filters, three blind spots

The results describe three filters, each with its own blind spot. The mandatory order sees crashes in volume, most of them minor, because telemetry tells operators about every contact. It cannot see operational failures without contact, even when they affect public safety, such as a robotaxi blocking an ambulance or a fleet freezing when traffic lights go dark. The media-derived database sees those operational failures and over-represents the most severe crashes, but it misses more than 99% of the crashes the regulator receives and filters out crashes in which the robot was not at fault. The regulatory register for surgical robots sees what the manufacturer decides to file; its volume can quadruple in a year without a comparable change in reported harm.

These findings fit the risk-regulation view that information gathering is shaped by triggers, reporters and filters (Hood et al., 2001). The crash trigger of the mandatory order defines the incident narrowly, which is the price of objectivity. The telemetry-based reporter overcomes the under-reporting that afflicts passive surveillance (Hazell & Shakir, 2006), but only for the events the trigger names. The rejection of H2 is consistent with the magnitude criterion of news selection (Galtung & Ruge, 1965; Harcup & O'Neill, 2017) and with the amplification of severe events (Kasperson et al., 1988); the unexpectedness criterion shows up instead in H1, through the large share of novel, harmless or near-harmless operational failures. The media channel therefore does not distort severity downwards; it compresses the distribution to its extremes, the most severe crashes and the most surprising behaviour.

For researchers, the implication is that a single register cannot stand in for the population of failures. Studies that pool media-derived incidents to characterise AI harm inherit both a severity filter and an attribution filter; studies that analyse

crash registers inherit a definition of incident that excludes operational failures; and trend analyses of device registers need to separate reporter behaviour from harm before they speak about safety.

6.2 The embodiment gap and its governance

Humanoid robots expose the gap most clearly. Their failures in 2026 involved loss of control, unwanted physical contact and alleged misuse in public spaces, the same classes of event that the crash order and the device register were built to capture in their own domains. Yet no register obliges anyone to report them, and the media-derived database records them late, if at all. The Collingridge (1980) dilemma applies with force: the period in which humanoid deployment can still be shaped is the period in which almost nothing about its failures is systematically recorded.

Four design implications follow, offered as options for technology assessment rather than as tested remedies. First, crash-triggered orders could add a category for operational failures without contact that affect public safety, such as obstruction of emergency services, immobilisation in live traffic or entry into closed areas; several such events already lead to recalls. Second, deployments of humanoid robots in public or commercial settings could carry a light notification duty for physical contact with people, loss of control and emergency stops, borrowing the trigger logic of the crash order: system engaged within a short window before the event, and faster deadlines for the most severe cases. Third, voluntary and mandatory registers could share identifiers so that researchers can link records without manual matching. Fourth, in the European Union, the serious-incident duty of the AI Act (Regulation (EU) 2024/1689, Article 73) and the Machinery Regulation (Regulation (EU) 2023/1230) leave open how general-purpose humanoids are classified; clarifying that scope would determine whether European humanoid incidents enter any register at all.

6.3 Limitations

The coding and linkage were performed by a single coder, with every decision documented for audit. Linkage relies on month-level incident dates and, for Level 2 vehicles, on redacted narratives, so two matches are probable rather than exact. Severity is unknown for most Level 2 incidents, which leaves most Tesla crashes outside the main contrast; treating them as non-severe strengthens the result (R3). With 18 linked incidents of known severity, interval estimates are wide, and the contrast under the third amended order alone rests on four links. The media-derived database is English-language and US-centric, and its indexing lags events, which affects the most recent humanoid cases. The regulatory decomposition uses problem codes, which co-occur and are truncated to the most frequent 100 per year. Domain identification by regular expressions may miss incidents described in unusual terms. Finally, none of the regimes records exposure, so the article says nothing about whether any system is safer or less safe than human operation.

7. Conclusion

This article compared what three incident regimes record about robotic AI failures between 2021 and 2026. Record-level linkage showed that the media-derived AI Incident Database contains almost every US crash it reports in the mandatory NHTSA register, but only a quarter of one per cent of the crashes the register holds. The pre-specified expectation that media coverage would favour harmless novelty over harm was not supported: visibility rises with severity, with an odds ratio of about nine between fatal or serious incidents and the rest. Novelty appears instead in a different class of failure: almost half of the US events in the media-derived database are operational failures that a crash trigger cannot capture. In the surgical register, a fourfold rise in malfunction reports traced back to two wear codes filed by the manufacturer, with little change in reported harm.

The contribution is a measured account of the embodiment gap: each regime sees a different slice of failure, the slices barely overlap, and the humanoid robots now entering public space fall outside all of them. For technology assessment, the lesson is that the safety of embodied AI cannot be read from any single register, and that the design of reporting

triggers deserves as much attention as the design of the robots themselves. Future work can extend the linkage to state-level registers and non-English incident collections, code device narratives to separate wear from failure, and test notification schemes for humanoid deployments before those deployments become routine.

Declarations

Data availability. The coding and linkage table, the NHTSA and openFDA raw files, the analysis script and all outputs are openly available on Zenodo at https://doi.org/10.5281/zenodo.DOI_PENDING (Merlet Echavarría, 2026). Because the AI Incident Database is licensed under CC BY-SA 4.0, the package is released under the same licence and does not redistribute the database export or the text of third-party reports; it includes a script that downloads the export of 12 October 2026 from the database's snapshot page.

Use of generative AI. Claude Opus 5.5 (Anthropic) was used under the author's direction to write the analysis code, process the data, search the literature, produce the figures, draft the text and make a first pass of the incident coding and record linkage. The author reviewed every coding and linkage decision, verified every reference against its primary source and takes full responsibility for the content.

Competing interests. The author receives publication incentives from Universidad del Alba and Instituto Profesional IACC. These institutions had no role in the design, analysis, interpretation or decision to publish.

Funding. This research received no specific grant from any funding agency.

Ethics. Not applicable. The study uses public records and published reports and involves no human participants.

Author contribution. Alfredo Agustín Merlet Echavarría: conceptualisation, methodology, data curation, formal analysis, validation, visualisation, writing (original draft, review and editing).

References

1. AI Incident Database (2026) AIID Excel export, 12 October 2026 [Data set]. Responsible AI Collaborative. <https://incidentdatabase.ai/research/snapshots/>
2. Alemzadeh H, Chen D, Lewis A, Kalbarczyk Z, Raman J, Leveson N, Iyer R (2015) Systems-theoretic safety assessment of robotic telesurgical systems. In *Computer safety, reliability, and security (Lecture Notes in Computer Science, Vol. 9337, pp. 213–227)*. Springer. https://doi.org/10.1007/978-3-319-24255-2_16
3. Alemzadeh H, Raman J, Leveson N, Kalbarczyk Z, Iyer RK (2016) Adverse events in robotic surgery: A retrospective study of 14 years of FDA data. *PLoS ONE* 11(4):e0151470. <https://doi.org/10.1371/journal.pone.0151470>
4. Andonian S, Okeke Z, Okeke DA, Rastinehad A, Vanderbrink BA, Richstone L, Lee BR (2008) Device failures associated with patient injuries during robot-assisted laparoscopic surgeries: A comprehensive review of FDA MAUDE database. *Can J Urol* 15(1):3912–3916
5. Armitage P (1955) Tests for linear trends in proportions and frequencies. *Biometrics* 11(3):375–386. <https://doi.org/10.2307/3001775>
6. Astle WA, Haus S (2026) Driving with influence: Exploring crash factors of automated systems in different roadway contexts (SAE Technical Paper 2026-01-0526). <https://doi.org/10.4271/2026-01-0526>. SAE International
7. CBS News (2026), October 14 German police seize humanoid robot that allegedly groped teen girl, and phones of young men who had it. <https://www.cbsnews.com/news/germany-robot-teen-girl-sexual-harassment-groping-cologne/>
8. Collingridge D (1980) *The social control of technology*. Frances Pinter

9. Correa J, Aribo C, Stuparich M, Nahas S, Cheung C, Behbehani S (2022) Malfunction events in the US FDA MAUDE database: How does robotic gynecologic surgery compare with other specialties? *J Minim Invasive Gynecol* 29(2):300–307e1. <https://doi.org/10.1016/j.jmig.2021.08.021>
10. eWeek (2026), February 19 China's Unitree aims to ship 20,000 humanoid robots in 2026. <https://www.eweek.com/news/unitree-20000-humanoid-robots-2026-china/>
11. Favarò FM, Nader N, Eurich SO, Tripp M, Varadaraju N (2017) Examining accident reports involving autonomous vehicles in California. *PLoS ONE* 12(9):e0184952. <https://doi.org/10.1371/journal.pone.0184952>
12. Friedman DCW, Lendvay TS, Hannaford B (2013) Instrument failures for the da Vinci surgical system: A Food and Drug Administration MAUDE database study. *Surg Endosc* 27(5):1503–1508. <https://doi.org/10.1007/s00464-012-2659-8>
13. Galtung J, Ruge MH (1965) The structure of foreign news. *J Peace Res* 2(1):64–90. <https://doi.org/10.1177/002234336500200104>
14. Global Times (2026) Robot malfunction at Shaanxi university event sparks debate on AI awareness and safety in human-robot interaction. <https://www.globaltimes.cn/page/202604/1359787.shtml>
15. Gupta P, Schomburg J, Krishna S, Adejoro O, Wang Q, Marsh B, Nguyen A, Genere JR, Self P, Lund E, Konety BR (2017) Development of a classification scheme for examining adverse events associated with medical devices, specifically the DaVinci surgical system as reported in the FDA MAUDE database. *J Endourol* 31(1):27–31. <https://doi.org/10.1089/end.2016.0396>
16. Hammadeh BM, Hussein AM, Alhamwi N, Mohammad AK, Ghanem R (2026) Device malfunctions and patient harm in robotic-assisted prostatectomy: A 17-year review of the FDA MAUDE database. *J Robotic Surg* 20(1):420. <https://doi.org/10.1007/s11701-026-03378-6>
17. Harcup T, O'Neill D (2017) What is news? News values revisited (again). *Journalism Stud* 18(12):1470–1488. <https://doi.org/10.1080/1461670X.2016.1150193>
18. Hazell L, Shakir SAW (2006) Under-reporting of adverse drug reactions: A systematic review. *Drug Saf* 29(5):385–396. <https://doi.org/10.2165/00002018-200629050-00003>
19. Hood C, Rothstein H, Baldwin R (2001) *The government of risk: Understanding risk regulation regimes*. Oxford University Press
20. Hsieh T-Y, Huang K-H (2026) Adverse events and device malfunctions in da Vinci robotic surgery: An analysis of MAUDE data from the recent decade. *J Robotic Surg* 20(1):72. <https://doi.org/10.1007/s11701-025-02947-5>
21. Kalra N, Paddock SM (2016) Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transp Res Part A: Policy Pract* 94:182–193. <https://doi.org/10.1016/j.tra.2016.09.010>
22. Kasperson RE, Renn O, Slovic P, Brown HS, Emel J, Goble R, Kasperson JX, Ratick S (1988) The social amplification of risk: A conceptual framework. *Risk Anal* 8(2):177–187. <https://doi.org/10.1111/j.1539-6924.1988.tb01168.x>
23. KTLA. (2026), September 3 The biggest fails from the 2026 World Humanoid Robot Games. <https://ktla.com/news/nationworld/world-humanoid-robot-games-fails/>
24. Kusano KD, Scanlon JM, Chen Y-H, McMurry TL, Chen R, Gode T, Victor T (2024) Comparison of Waymo rider-only crash data to human benchmarks at 7.1 million miles. *Traffic Inj Prev* 25(sup1):S66–S77. <https://doi.org/10.1080/15389588.2024.2380786>
25. Kusano KD, Scanlon JM, Chen Y-H, McMurry TL, Gode T, Victor T (2025) Comparison of Waymo rider-only crash rates by crash type to human benchmarks at 56.7 million miles. *Traffic Inj Prev* 26(sup1):S8–S20. <https://doi.org/10.1080/15389588.2025.2499887>

26. Kusano KD, Scanlon JM, McMurry TL, Victor T (2026) Do time of day and day of week matter when comparing Waymo's automated driving system and human crash involvement rates? *Traffic Injury Prevention*. <https://doi.org/10.1080/15389588.2026.2695421>. Advance online publication
27. Leveson NG (2011) *Engineering a safer world: Systems thinking applied to safety*. MIT Press
28. Lucas SM, Pattison EA, Sundaram CP (2012) Global robotic experience and the type of surgical system impact the types of robotic malfunctions and their clinical consequences: An FDA MAUDE review. *BJU Int* 109(8):1222–1227. <https://doi.org/10.1111/j.1464-410X.2011.10692.x>
29. Mantel N, Haenszel W (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22(4):719–748. <https://doi.org/10.1093/jnci/22.4.719>
30. May R, Krüger J, Leich T (2024) SoK: How artificial-intelligence incidents can jeopardize safety and security. In *Proceedings of the 19th International Conference on Availability, Reliability and Security (ARES 2024)* (Article 44). ACM. <https://doi.org/10.1145/3664476.3664510>
31. McGregor S (2021) Preventing repeated real world AI failures by cataloging incidents: The AI Incident Database. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17), 15458–15463. <https://doi.org/10.1609/aaai.v35i17.17817>
32. Medical Device Reporting, 21 C.F.R. pt. 803 (2026)
33. Merlet Echavarría AA (2026) The embodiment gap: Replication package [Data set]. Zenodo. https://doi.org/10.5281/zenodo.DOI_PENDING
34. Mimi MS, Islam MM, Barua S, Chowdhury TI, GhoshTusti A, Chhetri G, Das S (2026) Classifying crash severity by SAE automation level using transformer models: Insights for automated vehicle policy. *Transp Policy* 185:104222. <https://doi.org/10.1016/j.tranpol.2026.104222>
35. National Highway Traffic Safety Administration (2025) Third amended Standing General Order 2021-01: Incident reporting requirements. U.S. Department of Transportation. https://www.nhtsa.gov/sites/nhtsa.gov/files/2025-04/third-amended-SGO-2021-01_2025.pdf
36. National Highway Traffic Safety Administration (2026) Standing General Order on crash reporting: Incident report data (ADS and Level 2 ADAS; data through 15 September 2026) [Data set]. U.S. Department of Transportation. <https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting>
37. Nik-Ahd F, Souders CP, Houman J, Zhao H, Chughtai B, Anger JT (2019) Robotic urologic surgery: Trends in Food and Drug Administration-reported adverse events over the last decade. *J Endourol* 33(8):649–654. <https://doi.org/10.1089/end.2018.0802>
38. Paeth K, Atherton D, Pittaras N, Frase H, McGregor S (2025) Lessons for editors of AI incidents from the AI Incident Database. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28), 28946–28953. <https://doi.org/10.1609/aaai.v39i28.35163>
39. Patel D, Swaroop A, Bagchi S (2026) Investigating the relationship between autonomous driving system disengagements and crashes. In *Proceedings of the 56th Annual IEEE/IFIP International Conference on Dependable Systems and Networks: Supplemental Volume (DSN-S 2026)* (pp. 287–288). IEEE. <https://doi.org/10.1109/DSN-S70715.2026.00081>
40. Perrow C (1984) *Normal accidents: Living with high-risk technologies*. Basic Books
41. Power M (2007) *Organized uncertainty: Designing a world of risk management*. Oxford University Press
42. Raghupathi W, Ren J, Kulkarni T (2026) Examining the nature and dimensions of artificial intelligence incidents: A machine learning text analytics approach. *AppliedMath* 6(1):11. <https://doi.org/10.3390/appliedmath6010011>
43. Reason J (1997) *Managing the risks of organizational accidents*. Ashgate

44. Regulation (EU) (2023) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC and Council Directive 73/361/EEC. Official J Eur Union L 165:1–102
45. Regulation (EU) (2024) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689.
46. Shah MSB (2026) V2V and V2O crash injury analysis in ADAS and ADS vehicles using statistical and machine learning methods with explainable AI. In 2026 7th International Conference on Artificial Intelligence, Robotics, and Control (AIRC 2026) (pp. 414–420). IEEE. <https://doi.org/10.1109/AIRC69745.2026.11631431>
47. Tapper A, Leale D, Megahan G, Nacker K, Killinger K, Hafron J (2019) Robotic instrument failure: A critical analysis of cause and quality improvement strategies. *Urology* 131:125–129. <https://doi.org/10.1016/j.urology.2019.02.052>
48. Tuovinen L (2025) Death by AI: A survey of the literature and known incidents. *CEUR Workshop Proceedings*, 4237, 15–26
49. U.S. Food and Drug Administration (2026) openFDA device adverse event endpoint (MAUDE), product code NAY; data last updated 7 October 2026 [Data set]. <https://open.fda.gov/apis/device/event/>
50. Vaughan D (1996) *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. University of Chicago Press
51. Wang J, Harper CD, Hendrickson C (2026) Improving our understanding of contributing factors in California level 4 autonomous vehicle crashes through latent class analysis and machine learning-SHAP interpretations. *J Saf Res* 96:19–37. <https://doi.org/10.1016/j.jsr.2025.11.007>
52. Wei M, Zhou Z (2023) AI ethics issues in real world: Evidence from AI Incident Database. In *Proceedings of the 56th Hawaii International Conference on System Sciences* (pp. 4923–4932)
53. Wilson EB (1927) Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 22(158):209–212. <https://doi.org/10.1080/01621459.1927.10502953>
54. Yahoo Tech (2026) First humanoid robot 'arrested' after frightening a 70-year-old woman in China. <https://tech.yahoo.com/articles/first-humanoid-robot-arrested-frightening-144135993.html>

Figures

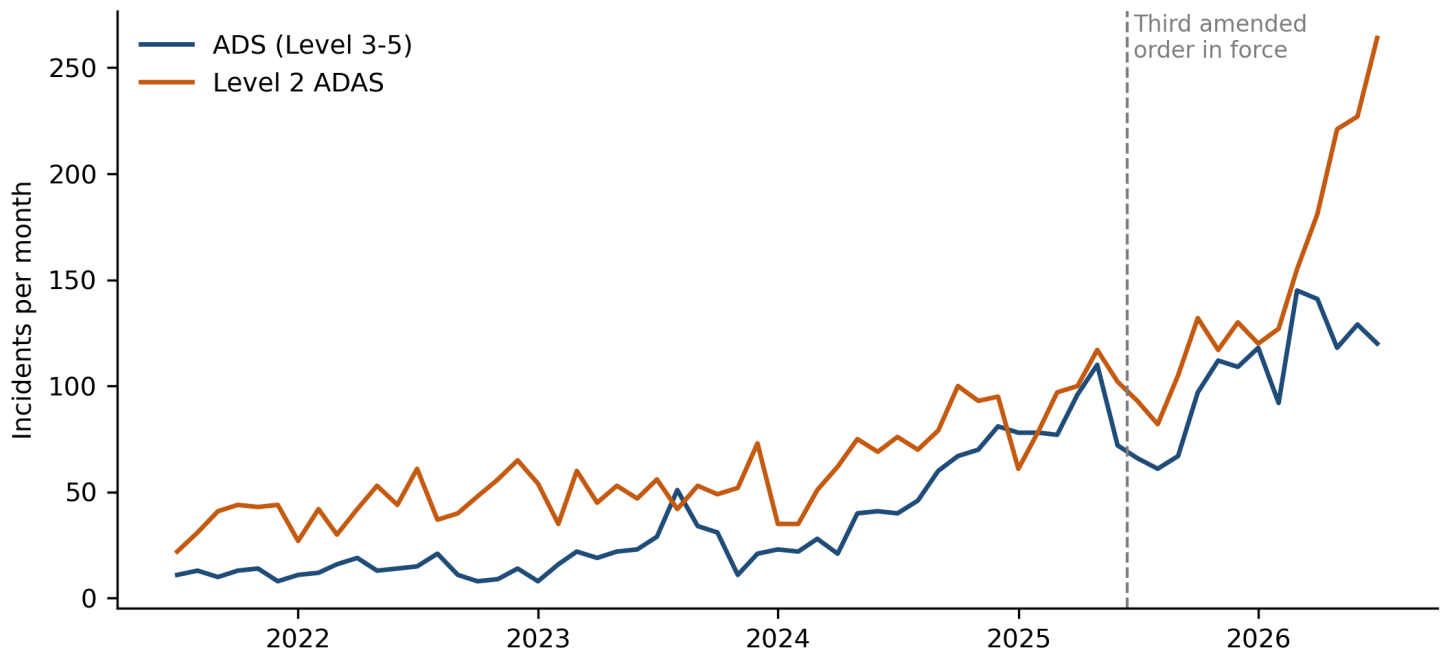


Figure 1

Incidents reported under NHTSA's Standing General Order by month of occurrence, July 2021 – July 2026

Note: incidents deduplicated by NHTSA's Same Incident ID; August and September 2026 omitted because reports for those months were still accruing. Source: author's elaboration from NHTSA (2026).

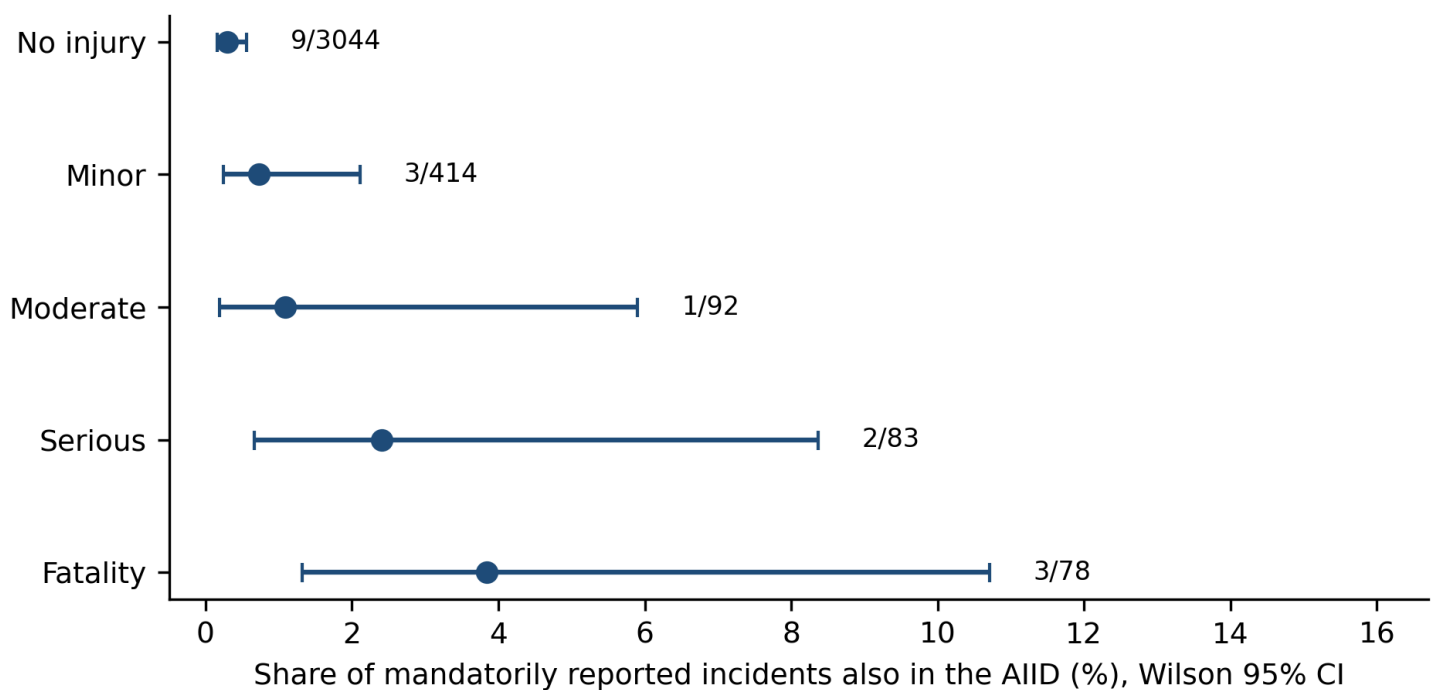


Figure 2

Share of mandatorily reported incidents that appear in the AI Incident Database, by severity

Note: points are shares; bars are Wilson 95% intervals; labels give linked incidents over all incidents at each level.
Source: author's calculations from NHTSA (2026) and AI Incident Database (2026).

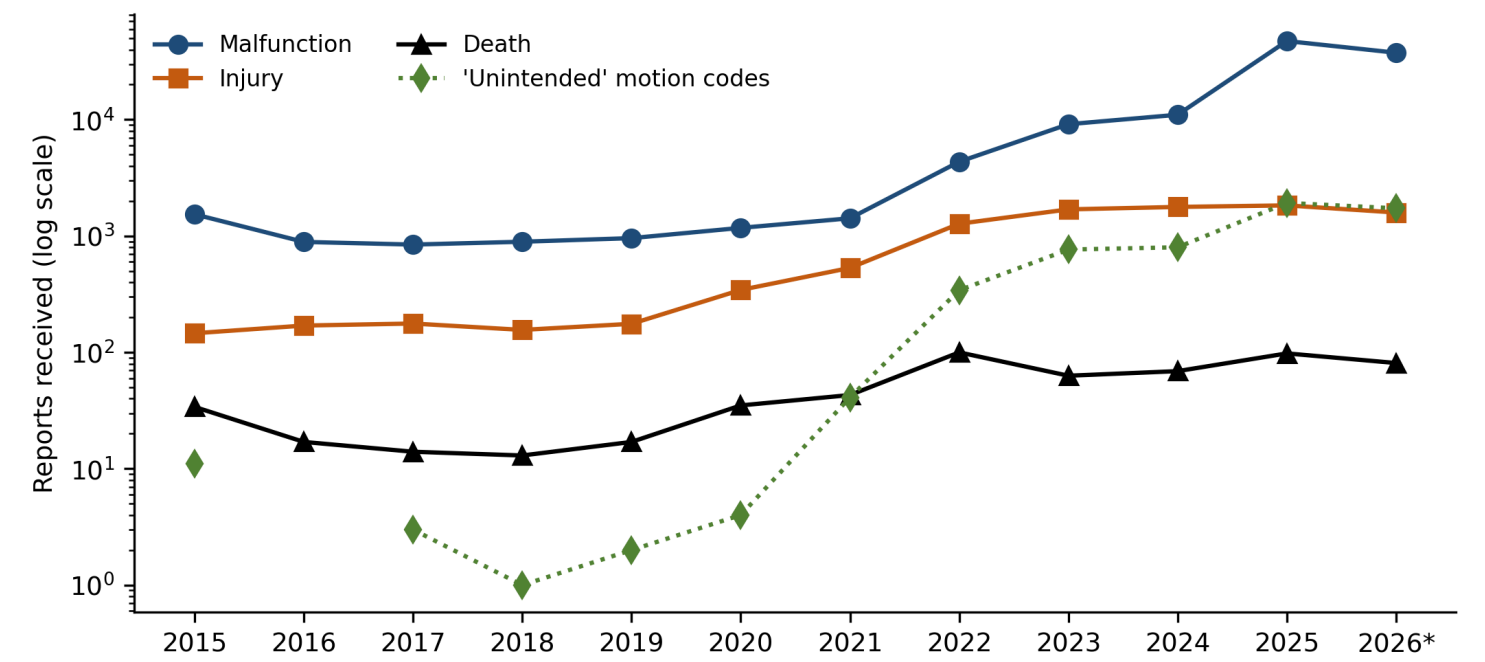


Figure 3

MAUDE reports for computer-assisted surgical systems (product code NAY) by event type, 2015–2026

Note: logarithmic scale; *2026 covers January–September; years with zero unintended-motion reports are not plotted.
Source: author's elaboration from U.S. Food and Drug Administration (2026).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixA.docx](#)