Physical AI Safety Maturity Model (PAS-MM): A Five-Level Framework for Industry Readiness

Mati Melchior
Independent Physical AI Safety Researcher
mati@physical-ai-safety.com

Version 1.0 · May 2026

Abstract

Physical AI deployments — humanoid robots, autonomous mobile robots, surgical systems, autonomous vehicles — are scaling faster than the safety vocabulary used to describe them. Safety claims across organizations are not comparable. Investors cannot benchmark. Regulators cannot set proportionate floors. Customers cannot procure on safety criteria. Adjacent fields solved this problem with maturity models: CMMI for software process, AI Safety Levels (ASL) for AI safety posture, ISO/IEC 33001 for process assessment, PCI DSS for payment-industry security. Physical AI lacks an equivalent.

This paper proposes the Physical AI Safety Maturity Model (PAS-MM), a five-level classification system: Ad-hoc, Documented, Compliant, Certified, and Defense-in-depth. Levels are anchored to recognized functional-safety standards (IEC 61508 SIL, ISO 13849 PL) and to the Physical AI Safety Stack introduced in a forthcoming companion paper.

PAS-MM includes a 30-question self-assessment instrument across six categories: hazard analysis, software safety, hardware safety, certification posture, operational monitoring, and incident response. A 0–90 score maps to a PAS Level via a defined rubric.

Application to approximately 30 publicly-known Physical AI organizations, using public information only, shows clustering at PAS 1–2, a small population at PAS 3–4, and no organization at PAS 5. The paper concludes with adoption guidance for analysts, regulators, insurers, customers, and Physical AI organizations.

Keywords: Physical AI Safety, maturity model, IEC 61508, ISO 13849, certification, industry assessment.

---

1 Introduction

Physical AI is the class of AI systems that act on the physical world through actuators — humanoid robots, autonomous mobile robots, surgical systems, autonomous vehicles, drones, and similar platforms. The term and its scoping are introduced in a forthcoming companion paper on the Physical AI Safety definitional framework. Deployment volumes are rising. Per the Israel Innovation Authority's April 2026 AI Strategy report and the IFR *World Robotics 2025 — Industrial Robots* report (released September 2025), the global operational stock of industrial robots reached 4.664 million units in 2024, with 542,000 new installations that year — the second-highest annual installation count in history. The AI-driven subset of this fleet is growing but is not separately reported.

Safety practice across this sector is uneven. Some organizations certify safety functions to IEC 61508 and ISO 13849. Others rely on software watchdogs and best-effort testing. Public safety claims are written in incompatible vocabularies. The terms "safe," "safety-rated," "safety-certified," and "safety-aware" appear in marketing material with no shared meaning.

This raises four operational problems:

- **Investors** cannot compare safety posture across portfolio candidates.
- **Regulators** cannot set proportionate compliance floors.
- **Customers** cannot include safety as a procurement criterion.
- **Insurers** cannot price risk consistently across this sector.

Adjacent fields solved this problem with maturity models. The Capability Maturity Model Integration (CMMI), maintained by the Carnegie Mellon Software Engineering Institute, has been operational for thirty years and shaped software-process discipline globally. Anthropic's AI Safety Levels (ASL) framework, introduced in 2023, has rapidly become a reference vocabulary in the AI alignment field. ISO/IEC 33001 underlies process-assessment standards in many domains. PCI DSS levels organize compliance posture in the payment industry.

Physical AI lacks an equivalent framework.

This paper proposes the Physical AI Safety Maturity Model (PAS-MM): a five-level classification anchored to existing functional-safety standards, with a 30-question self-assessment instrument and an industry-distribution baseline.

The remainder of the paper is organized as follows. Section 2 reviews maturity-model precedents in adjacent fields. Section 3 defines the five PAS Levels. Section 4 presents the self-assessment instrument and scoring rubric. Section 5 applies PAS-MM to approximately 30 publicly-known organizations. Section 6 derives implications for stakeholders. Section 7 outlines the roadmap from PAS 1 to PAS 5. Section 8 concludes.

---

2 Background: maturity models in adjacent fields

A maturity model defines progressive levels of organizational capability against a domain. Each level specifies measurable attributes. Organizations self-assess, are externally assessed, or are aggregated by analysts. Maturity models become operational when three or more independent stakeholder classes adopt them: organizations (for self-positioning), customers and analysts (for benchmarking), and regulators (for floors).

2.1 CMMI

CMMI defines five process-maturity levels: Initial, Managed, Defined, Quantitatively Managed, and Optimizing. The framework has been maintained by the Carnegie Mellon Software Engineering Institute since the 1990s and operates today as CMMI v2.0. Industry-wide adoption shaped procurement, government contracting, and process discipline in large software organizations for three decades.

2.2 AI Safety Levels (ASL)

Anthropic introduced the AI Safety Levels framework in 2023 as part of its Responsible Scaling Policy. ASL defines a sequence of levels (ASL-1 through ASL-5+) based on the catastrophic-risk potential of AI systems and the corresponding safety practices required. ASL has rapidly become a reference framework cited in AI policy, standards work, and competing labs' published commitments. The framework shows that a maturity-style model in an adjacent emerging field can become operational within months when the underlying gap is acute.

2.3 ISO/IEC 33001

The ISO/IEC 33000 series, beginning with ISO/IEC 33001, defines process-assessment standards used as the foundation for many domain-specific maturity frameworks. The series specifies how to define a process-assessment model, how to score against it, and how to report results.

2.4 PCI DSS

The Payment Card Industry Data Security Standard (PCI DSS) defines four merchant levels based on transaction volume, with progressively stricter compliance requirements. PCI DSS shows that a sector-specific maturity framework can become procurement-mandatory: a merchant that cannot show its level cannot process card transactions.

2.5 Functional safety as anchor: IEC 61508 and ISO 13849

PAS-MM differs from CMMI and ASL in one material way: Physical AI safety has existing, mature anchor standards. IEC 61508 defines Safety Integrity Levels (SIL 1–4) for electrical, electronic, and programmable electronic safety-related systems. ISO 13849-1:2023 defines Performance Levels (PL a–e) for safety-related parts of machinery control systems. ISO/IEC TR 5469 addresses functional safety and AI systems explicitly. Recurring-authority publications such as the Air Street State of AI report show that annual industry frameworks can become reference points.

PAS-MM uses these standards as the certification anchor for higher levels rather than competing with them.

2.6 Operational threshold

In each precedent above, the framework became operational once three or more stakeholder classes adopted it. This three-stakeholder threshold informs the adoption argument in Sections 6 and 8.

---

3 The five PAS Levels

The **Physical AI Safety Maturity Model (PAS-MM)** is a five-level classification system for Physical AI organizations based on the visibility, depth, and certification posture of their safety infrastructure. It is intended for self-assessment, third-party assessment, and industry-wide benchmarking. PAS-MM is functionally analogous to CMMI for software process maturity and to AI Safety Levels (ASL) for AI safety posture.

The five levels rise from PAS 1 (Ad-hoc) at the base to PAS 5 (Defense-in-depth) at the top (see Figure 1).

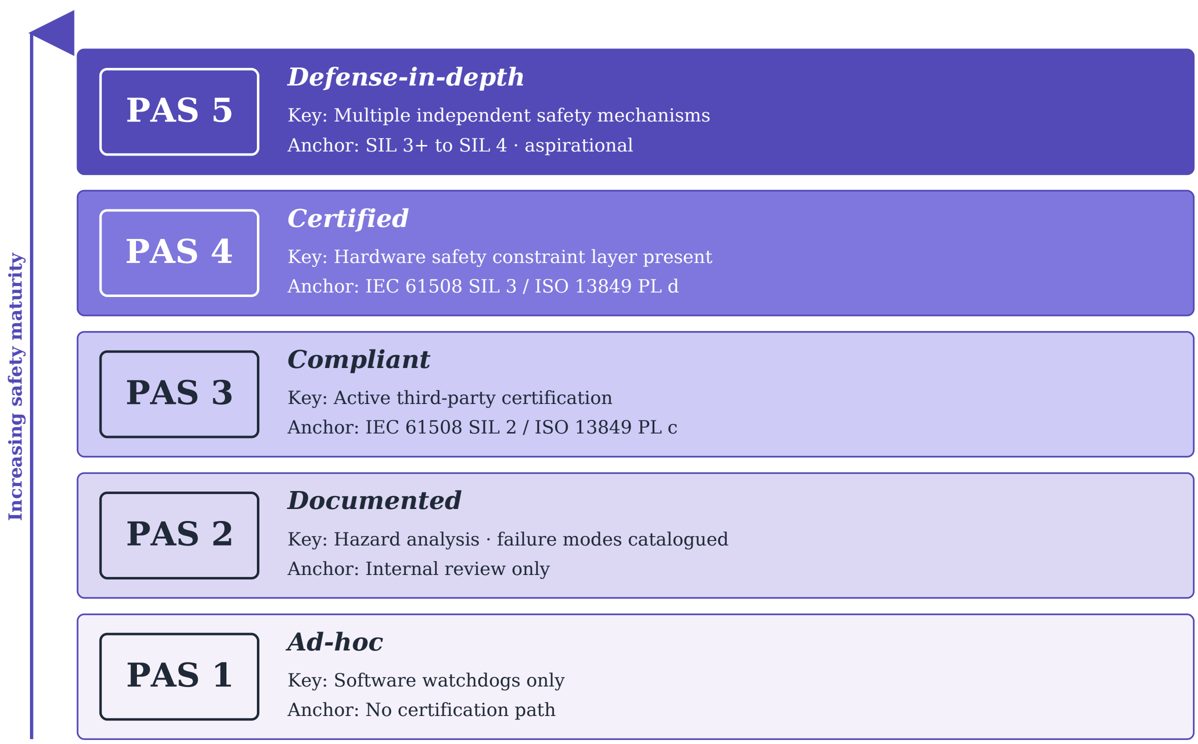

Figure 1. Vertical five-level PAS-MM ladder. From base to top: PAS 1 Ad-hoc with software watchdogs and no certification path; PAS 2 Documented with hazard analysis and internal review only; PAS 3 Compliant with active third-party certification anchored to IEC 61508 SIL 2 or ISO 13849 PL c; PAS 4 Certified with hardware safety constraint layer anchored to SIL 3 or PL d; PAS 5 Defense-in-depth with multiple independent safety mechanisms targeting SIL 3+ to SIL 4. Single-hue gradient from light at PAS 1 to dark at PAS 5.

Figure 1. PAS Level ladder. Vertical five-level visualization of PAS-MM. PAS 1 (Ad-hoc) at the base; PAS 5 (Defense-in-depth) at the top. Each level is annotated with name, key characteristic, and example certification anchor. The single-hue gradient (light to dark) tracks increasing safety maturity.

Each level is defined below.

3.1 PAS Level 1 — Ad-hoc

> **PAS Level 1 — Ad-hoc.** No documented safety architecture. Software watchdogs only. Safety is "we'll catch it in testing." No certification path. No third-party review. No documented hazard analysis. Most early-stage humanoid and autonomous-mobile-robot startups are PAS 1.

Characteristic indicators. No public safety case. No published hazard analysis. Marketing references "safety" without specifying a standard. Engineering teams cite testing coverage but not formal verification, failure-mode analysis, or independent review.

Failure modes addressed. Common single-point software faults that surface during functional testing.

Failure modes not addressed. Hardware faults, common-cause failures, systematic faults in the AI stack, faults discovered only in field deployment, and any failure mode requiring a documented mitigation chain.

Typical organizations. Pre-seed and seed-stage humanoid and autonomous mobile robot startups; many academic robot platforms when transitioning to commercial use.

3.2 PAS Level 2 — Documented

> **PAS Level 2 — Documented.** Safety requirements documented (hazard analysis, failure modes catalogued). Software-based monitoring and run-time assurance present. Internal review only. No third-party certification. Pre-Series-A companies often plateau here.

Characteristic indicators. Internal hazard-analysis document exists. Failure modes are catalogued, often in spreadsheet form. Software safety monitors are present in the runtime stack. Internal safety reviews occur but are not externally audited. Safety claims in marketing material reference the existence of these practices without citing a standard.

Failure modes addressed. Documented software faults with known mitigations; faults observed during structured internal testing.

Failure modes not addressed. Faults requiring third-party verification; faults that depend on hardware enforcement; faults outside the imagination of the internal team.

Typical organizations. Pre-Series-A and early Series-A Physical AI companies, particularly those that have hired their first dedicated safety engineer but have not yet engaged a notified body.

3.3 PAS Level 3 — Compliant

> **PAS Level 3 — Compliant.** Active third-party certification effort to a recognized functional safety standard (IEC 61508 SIL 2 or ISO 13849 PL c, or equivalent). Mixed software/hardware safety mechanisms. Documented safety case. Established Physical AI companies aim for PAS 3 as deployment minimum for human-adjacent applications.

Characteristic indicators. Engagement with a notified body or accredited certifying entity. Documented safety case with traceability from hazards to mitigations to verification evidence. Mixed software and hardware safety mechanisms in the runtime stack. Public statements identifying the targeted standard (IEC 61508 SIL 2, ISO 13849 PL c, or equivalent).

Failure modes addressed. Faults within the scope of the targeted SIL or PL, with mitigations verified by third-party review.

Failure modes not addressed. Faults requiring SIL 3+ integrity (the gap between random-hardware-failure rates required at SIL 2 versus SIL 3 is approximately two orders of magnitude); faults arising from common-cause failures across mixed software-hardware mechanisms.

Typical organizations. Established industrial robot OEMs for legacy product lines; well-funded Series-B and later Physical AI companies in regulated deployment contexts (logistics, hospital, defense-adjacent).

3.4 PAS Level 4 — Certified

> **PAS Level 4 — Certified.** SIL 3 / PL d certification achieved or mature in-progress. Hardware safety constraint layer (Layer 3 of the Physical AI Safety Stack from a forthcoming companion paper on the Physical AI Safety Stack) present and verified. Independent third-party audit. Published safety case. Industrial robot OEMs and well-funded humanoid startups target PAS 4 for safety-critical applications.

Rationale for the hardware layer requirement. PAS 4 introduces a hardware safety constraint layer because software safety alone reaches a ceiling at the SIL 2 to SIL 3 boundary. The argument that systematic and common-cause failure rates in software-only safety mechanisms exceed the per-hour failure budget required at higher SIL levels is the subject of a forthcoming companion paper on the software safety ceiling. Per IEC 61508-1:2010 Table 3, high-demand mode SIL 3 corresponds to a probability of failure per hour of 10^{-7} to 10^{-8}. Hardware safety constraint layers — Layer 3 of the Physical AI Safety Stack — are independent of the AI compute path and are verifiable against the higher integrity targets.

Characteristic indicators. SIL 3 or PL d certification achieved or with documented progress milestones. Independent third-party audit performed and reported. Published safety case (full or summary). Hardware safety mechanisms physically and logically separable from the AI compute path.

Failure modes addressed. Random hardware faults at SIL 3 rates (10^{-7} to 10^{-8}

per hour for high-demand mode, per IEC 61508-1:2010 Table 3); systematic faults addressed via the safety case process; common-cause faults across software-hardware boundaries when the hardware layer is verified independent.

Failure modes not addressed. Faults requiring multiple independent safety mechanisms (a PAS 5 condition); faults arising from operational data not fed back into the safety case.

Typical organizations. Major industrial robot OEMs for safety-rated product variants; well-funded humanoid startups with dedicated safety hardware programs; regulated autonomous vehicle programs at the certification stage.

3.5 PAS Level 5 — Defense-in-depth

> **PAS Level 5 — Defense-in-depth.** SIL 3+ to SIL 4 certified. Multiple independent safety mechanisms. Hardware safety constraint layer with verified separation of fault domains. Continuous third-party monitoring. Published incident response capability. Operational data shared with regulators. As of publication, no commercially deployed Physical AI system is at PAS 5; it is the aspirational target equivalent to aviation / nuclear / rail safety standards in those domains.

Characteristic indicators. SIL 3+ or SIL 4 certification with multiple independent safety mechanisms. Verified separation of fault domains across the safety architecture. Continuous monitoring by an independent entity. Published incident response capability with disclosed timelines. Operational safety data shared with regulators or industry consortia.

Failure modes addressed. Beyond SIL 3 — the regime of aviation, nuclear, and rail safety practice.

Aspirational status. As of publication, no commercially deployed Physical AI system meets these criteria. PAS 5 sets the long-horizon target.

3.6 Mapping to adjacent maturity frameworks

PAS-MM is positioned as complementary to CMMI, ASL, and IEC 61508. Figure 2 presents the cross-framework mapping.

Figure 2. Mapping of PAS Levels to adjacent maturity and integrity frameworks. Approximate mappings — frameworks measure different attributes.

The mapping is approximate because the source frameworks measure different attributes: CMMI measures process maturity, ASL measures model risk class, and IEC 61508 SIL measures dangerous-failure rate. PAS-MM aggregates these into an organizational safety-posture classification specific to Physical AI.

PAS Level	CMMI	ASL	IEC 61508 SIL
PAS 5	Optimizing	ASL-5 (aspirational)	SIL 3+ to SIL 4
PAS 4	Quantitatively Managed	ASL-4	SIL 3
PAS 3	Defined	ASL-3	SIL 2
PAS 2	Managed	ASL-2	SIL 1
PAS 1	Initial	ASL-1	(none / SIL 1 partial)

Approximate mappings — frameworks measure different attributes.

Figure 2. Five-row by four-column mapping matrix. Columns: PAS Level, CMMI, ASL, IEC 61508 SIL. Rows from top to bottom: PAS 5 maps to CMMI Optimizing, ASL-5 (aspirational), SIL 3+ to SIL 4. PAS 4 maps to CMMI Quantitatively Managed, ASL-4, SIL 3. PAS 3 maps to CMMI Defined, ASL-3, SIL 2. PAS 2 maps to CMMI Managed, ASL-2, SIL 1. PAS 1 maps to CMMI Initial, ASL-1, none or SIL 1 partial. Row colors follow the maturity gradient from light at PAS 1 to dark at PAS 5.

4 Self-assessment instrument

The **PAS-MM Self-assessment Instrument** is a structured questionnaire across six categories: hazard analysis, software safety, hardware safety, certification posture, operational monitoring, and incident response. Each category has 5 questions; total of 30 questions. Each question is scored 0–3. The total score (0–90) maps to a PAS Level via a defined rubric.

Scoring scale per question:

- **0** — Not performed or not documented.
- **1** — Performed informally; no documented evidence.
- **2** — Performed and documented; internal review only.
- **3** — Performed, documented, and externally verified.

The six categories follow.

4.1 Category 1 — Hazard analysis

#	Question (0/1/2/3)
Q1	Has the organization performed a documented hazard analysis?
Q2	Is the hazard analysis updated when product changes?
Q3	Are failure modes documented and traceable to mitigations?

#	Question (0/1/2/3)
Q4	Is the hazard analysis reviewed by a third party?
Q5	Is the hazard analysis publicly available (or available under NDA to customers)?

4.2 Category 2 — Software safety

#	Question (0/1/2/3)
Q6	Are software safety requirements derived from the hazard analysis?
Q7	Is software safety verified against requirements?
Q8	Is software safety tested under failure injection?
Q9	Is software safety verified by formal methods or independent V&V?
Q10	Is software safety certified to a recognized standard?

4.3 Category 3 — Hardware safety

This category references Layer 3 of the Physical AI Safety Stack as introduced in a forthcoming companion paper on the Physical AI Safety definitional framework.

#	Question (0/1/2/3)
Q11	Is there a documented hardware safety layer (Layer 3 of the Stack)?
Q12	Is the hardware safety layer verified independently of the AI compute?
Q13	Is separation of fault domains documented and verified?
Q14	Is the hardware safety layer certified to IEC 61508-2 or equivalent?
Q15	Is the hardware safety mechanism verified against common-cause failure?

4.4 Category 4 — Certification posture

#	Question (0/1/2/3)
Q16	Does the organization engage with notified bodies / certifying entities?

#	Question (0/1/2/3)
Q17	Is the safety case documented and reviewed?
Q18	Is the certification scope appropriate for the deployment context?
Q19	Is certification renewed / maintained over product lifecycle?
Q20	Are certification artifacts public or available under NDA to customers?

4.5 Category 5 — Operational monitoring

#	Question (0/1/2/3)
Q21	Are safety-relevant operational events logged?
Q22	Are logs immutable and tamper-evident?
Q23	Is there continuous monitoring of safety performance?
Q24	Are anomalies reported to relevant stakeholders?
Q25	Are operational data shared with regulators or industry consortia?

4.6 Category 6 — Incident response

#	Question (0/1/2/3)
Q26	Is there a documented incident response procedure?
Q27	Has the procedure been tested in drills or actual incidents?
Q28	Are root-cause analyses performed and acted on?
Q29	Are lessons learned shared internally?
Q30	Are lessons learned shared externally (within reasonable bounds)?

4.7 Scoring rubric

The total score (0–90) maps to a PAS Level as follows.

Table 1. Scoring rubric mapping total score to PAS Level.

Total score	PAS Level
0–25	PAS Level 1
26–45	PAS Level 2
46–65	PAS Level 3
66–80	PAS Level 4
81–90	PAS Level 5

Self-assessment is intended to take less than two hours for an informed respondent. The instrument is provided as supplementary material and may be reproduced under the CC-BY 4.0 license of this paper.

---

5 Industry application

Industry application of PAS-MM uses publicly available information to classify Physical AI organizations. Classifications using public information only should be marked with a confidence indicator (high / medium / low). Self-classifications by organizations are accepted at face value and noted as "self-classified."

5.1 Methodology

The application reported here covers approximately 30 publicly-known Physical AI organizations across five sub-sectors: legacy industrial robot OEMs, surgical robotics manufacturers, defense and aerospace robotics, AI-native humanoid and autonomous mobile robot startups, and logistics and warehouse automation.

Classification used the following sources only: regulatory filings, public certification records, published safety cases, peer-reviewed publications, technical conference presentations, and corporate communications. Classifications based primarily on regulatory filings or public certification records were marked **high** confidence. Classifications based on corporate communications or analyst coverage were marked **medium** confidence. Classifications based on absence of public information (interpreted as "no evidence of higher posture") were marked **low** confidence.

Where public evidence was insufficient for any classification with at least low confidence, the organization was excluded. Where classification was possible but identifying the organization could create disproportionate exposure to either party, the classification is reported in aggregate without naming.

Organizations are invited to contact the author with corrections or self-classifications. Estimates use public information; companies are invited to contact the author with corrections or self-classifications. Self-classifications received before the next revision will be reflected in the next published version of this analysis.

A forthcoming companion paper applies the same methodology at industry scale and provides per-organization detail at the appropriate confidence levels.

5.2 Distribution

The distribution across PAS Levels is shown in Figure 3.

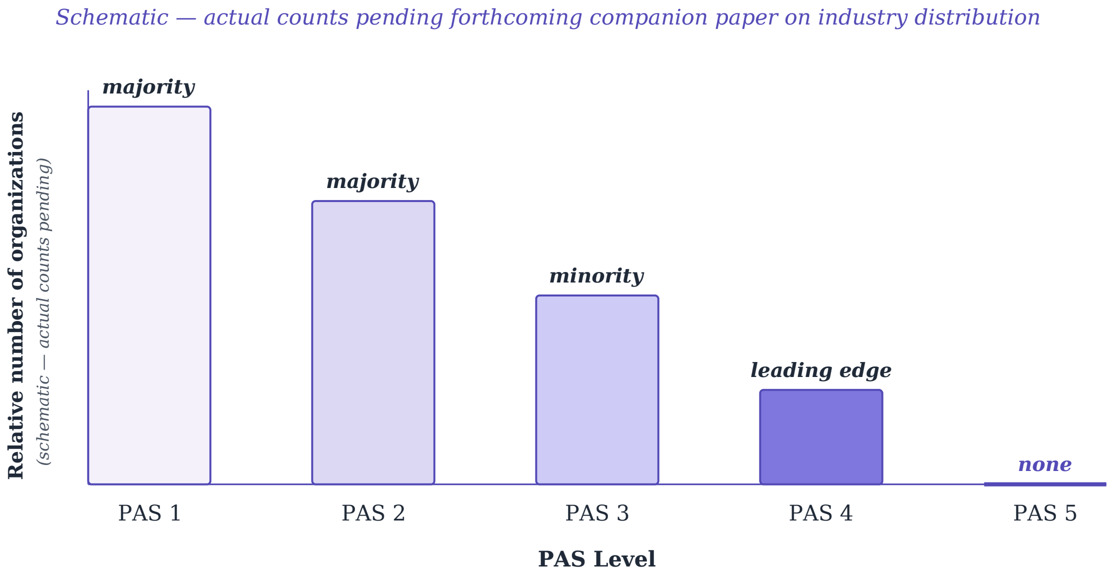

Figure 3. Schematic histogram showing the right-skewed distribution of approximately 30 publicly-known Physical AI organizations across the five PAS Levels. Bar heights are relative and show the qualitative shape only — actual counts are pending the forthcoming companion paper on industry distribution. Most organizations cluster at PAS 1 and PAS 2; a minority is at PAS 3; a small leading edge is at PAS 4; no organization is at PAS 5 as of mid-2026. Bar colors follow the locked maturity gradient from light at PAS 1 to dark at PAS 5.

Figure 3. Schematic distribution of publicly-known Physical AI organizations across PAS Levels. Bar heights are relative; actual counts are pending the forthcoming companion paper on industry distribution.

Distribution across PAS Levels in publicly observable Physical AI organizations is right-skewed: a majority cluster at PAS Level 1–2 (ad-hoc to documented but uncertified), a minority at PAS Level 3 (active certification effort), a small leading edge at PAS Level 4 (achieved certification), and as of mid-2026 no commercially deployed Physical AI system is at PAS Level 5.

The majority of surveyed organizations classify at PAS 1 or PAS 2.

Confidence is **high** for legacy industrial OEMs in the PAS 3–4 cluster, where ISO 13849 and IEC 61508 certification records are public. Confidence is **medium** for surgical robotics organizations in the PAS 3 cluster, where FDA Class II/III clearance establishes a baseline that maps approximately to PAS 3. Confidence is **medium-to-low** for AI-native startups in the PAS 1–2 cluster, where the classification reflects the absence of public certification evidence rather than evidence against certification.

5.3 Common patterns observed

Five patterns emerge from the distribution:

1. **Legacy industrial robot OEMs** (such as ABB, KUKA, Fanuc, Yaskawa, and DENSO) tend to PAS 3–4 for legacy product lines, consistent with decades of ISO 13849 and IEC 61508 practice. Confidence: **high**, based on public certification records.
2. **AI-native startups** in the humanoid and AMR categories tend to PAS 1–2. The dominant safety mechanism is software watchdogs combined with operational restrictions (cages, low speed, low payload). Confidence: **medium-to-low**, based on absence of public certification evidence.
3. **Defense and aerospace Physical AI** programs tend to operate at PAS 3 or higher, driven by parallel adjacent regulatory regimes (military airworthiness, defense safety standards). Confidence: **medium**, based on regulatory regime adjacency.
4. **Surgical robotics** organizations tend to map to FDA Class II or Class III clearance, which corresponds approximately to PAS 3 to PAS 4 in PAS-MM terms, reflecting the medical-device regulatory baseline. Confidence: **medium**, based on public clearance records.
5. **Humanoid and autonomous mobile robot startups** with public-facing demos cluster at PAS 1–2, with a small leading edge approaching PAS 3 through publicly declared certification roadmaps. Confidence: **medium-to-low**, based on corporate communications.

---

6 Implications for stakeholders

PAS-MM is intended to be used. Six stakeholder classes have direct application.

Stakeholder	Application of PAS Level
Investors	Due diligence input. PAS 1–2 startups carry meaningful tail risk in human-adjacent deployment contexts.
Regulators	Mandating PAS 3 minimum for human-adjacent deployments would establish a sectoral floor analogous to ISO 13849 PL c.
Insurers	PAS Level should map to underwriting tiers. Below PAS 3, insurance should be qualified or tier-priced.
Customers (industrial buyers, hospitals, fleets)	Procurement criteria should include PAS Level for safety-critical applications.

Stakeholder	Application of PAS Level
Standards bodies	PAS-MM is a starting point for formal safety integrity standards specific to Physical AI.
Compliance officers	Internal benchmark and target-setting for cross-product safety posture.

These applications are not mutually exclusive. The framework becomes operational once three or more of these stakeholder classes adopt it, in line with the precedents reviewed in Section 2.

Adoption rarely begins with all six stakeholders simultaneously. The CMMI and ASL precedents show a typical sequence: organizations self-position first, customers and analysts adopt the vocabulary second, and regulators and insurers arrive third. PAS-MM is structured to support that sequence. The self-assessment instrument enables stage-one self-positioning at low cost. The public mapping in Section 5 supports stage-two analyst adoption. The IEC 61508 and ISO 13849 anchors give stage-three regulators and insurers an existing standards basis to reference, lowering the cost of regulatory and underwriting integration.

---

7 Roadmap from PAS 1 to PAS 5

The following roadmap describes typical progression. Calendar time varies with funding, regulatory context, and product complexity. Representative end-to-end progression from PAS 1 to PAS 5 takes 5 to 10 years for a Physical AI organization.

PAS 1 → PAS 2. Document everything. Perform and write up the hazard analysis. Catalogue failure modes. Implement and document software safety mechanisms. Establish an internal safety review process.

PAS 2 → PAS 3. Engage a third-party certifier. Begin a formal safety case. Plan for IEC 61508 SIL 2 or ISO 13849 PL c certification of at least one safety function. Establish traceability from hazards to mitigations to verification evidence.

PAS 3 → PAS 4. Add a hardware safety constraint layer (Layer 3 of the Stack). Plan for SIL 3 or PL d certification. Commission an independent third-party audit. Publish the safety case.

PAS 4 → PAS 5. Implement multiple independent safety mechanisms. Verify separation of fault domains across the safety architecture. Establish continuous third-party monitoring. Publish incident response capability and share operational data with regulators or industry consortia.

Typical effort estimates: each level transition requires on the order of months for early-stage transitions (PAS 1→2), one to two years for mid-stage (PAS 2→3), two to three years for advanced (PAS 3→4), and multi-year sustained programs for the highest level (PAS 4→5). The full PAS 1→5 trajectory typically spans five to ten years for a Physical AI organization.

The four transitions are visualized as a staircase in Figure 4.

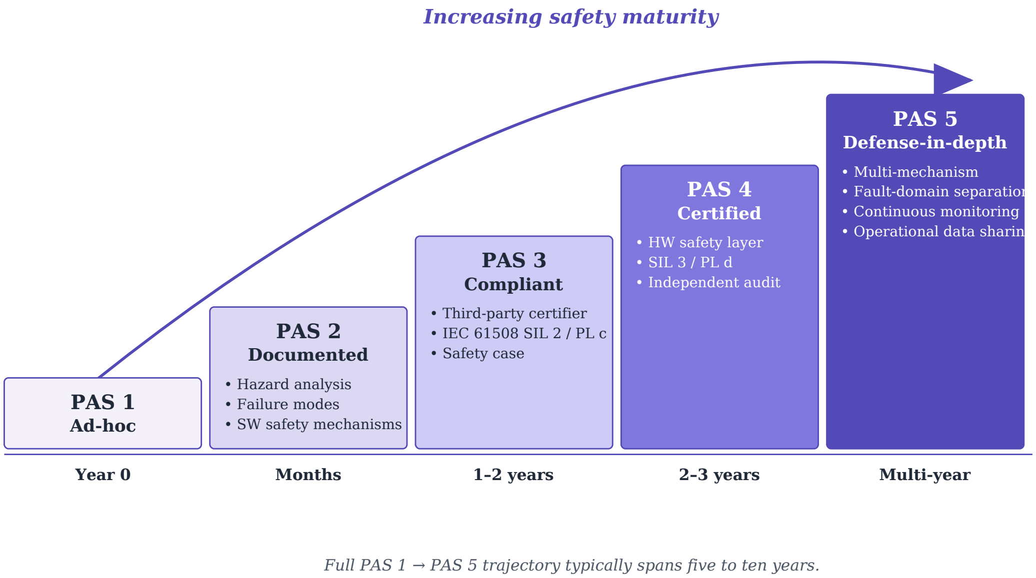

Figure 4. Staircase diagram of the PAS-MM roadmap with five ascending steps from PAS 1 to PAS 5. PAS 1 Ad-hoc at Year 0 with software watchdogs and no certification path. PAS 2 Documented at months scale with hazard analysis, failure modes, and software safety mechanisms. PAS 3 Compliant at one to two years with third-party certifier engagement, IEC 61508 SIL 2 or PL c targets, and a documented safety case. PAS 4 Certified at two to three years with hardware safety layer, SIL 3 or PL d certification, and independent audit. PAS 5 Defense-in-depth at multi-year scale with multi-mechanism architecture, fault-domain separation, continuous monitoring, and operational data sharing. An arc arrow rising from PAS 1 to PAS 5 is labeled increasing safety maturity. Step colors follow the maturity gradient from light at PAS 1 to dark at PAS 5.

Figure 4. Roadmap from PAS 1 to PAS 5. Staircase diagram showing the four transitions, the key actions at each step, and approximate calendar time per stage.

The estimate of 5 to 10 years is consistent with observed industrial precedent: legacy industrial robot OEMs took comparable time to mature their functional-safety practices to current PAS 3–4 levels. Forthcoming companion papers on reference architecture, common-cause failure analysis, regulatory convergence, and the certification framework provide additional detail on each transition.

---

8 Conclusion

This paper proposes the Physical AI Safety Maturity Model (PAS-MM): a five-level classification system for Physical AI organizations, anchored to recognized functional-safety standards, supported by a 30-question self-assessment instrument, and applied to approximately 30 publicly-known organizations.

The framework follows the operational pattern established by CMMI, ASL, ISO/IEC

33001, and PCI DSS. Each of these became operational once three or more independent stakeholder classes adopted them. PAS-MM is offered to that adoption process as a starting point, not a final word.

We invite Physical AI organizations to self-classify, regulators to consider PAS 3 as a deployment floor for human-adjacent applications, insurers to align underwriting with PAS Levels, customers to include PAS Level in procurement criteria, analysts to use the framework in sector reports, and standards bodies to refine and extend it. Annual revisions are intended.

Industry maturity is measured. What gets measured, gets managed.

References

[1] CMMI Institute / Carnegie Mellon University Software Engineering Institute, "CMMI v2.0 Model," Pittsburgh, PA, USA.

[2] Anthropic, "Anthropic's Responsible Scaling Policy" (AI Safety Levels framework), 2023.

[3] ISO/IEC 33001:2015, "Information technology — Process assessment — Concepts and terminology," International Organization for Standardization / International Electrotechnical Commission.

[4] IEC 61508-1:2010, "Functional safety of electrical / electronic / programmable electronic safety-related systems — Part 1: General requirements," International Electrotechnical Commission.

[5] ISO 13849-1:2023, "Safety of machinery — Safety-related parts of control systems — Part 1: General principles for design," International Organization for Standardization.

[6] ISO/IEC TR 5469, "Artificial intelligence — Functional safety and AI systems," International Organization for Standardization / International Electrotechnical Commission.

[7] PCI Security Standards Council, "Payment Card Industry Data Security Standard (PCI DSS) — Levels Documentation."

[8] Israel Innovation Authority, "AI Strategy Report," April 2026.

[9a] International Federation of Robotics, "World Robotics 2025 — Industrial Robots," released 25 September 2025. Available: ifr.org/worldrobotics/

[9b] International Federation of Robotics, "World Robotics 2025 — Service Robots," released 7 October 2025. Available: ifr.org/worldrobotics/

[10] Air Street Capital, "State of AI Report."

Mati Melchior · Independent Physical AI Safety Researcher mati@physical-ai-safety.com · physical-ai-safety.com P4 of an 8-paper series on Physical AI Safety.