Don’t Call Me Artificial

The STAMP Methodology for Creating Standardized System Profiles for AI Implementations

Paul Holcomb

Version 09.27.25

Executive Summary:

This paper introduces **STAMP (System Taxonomy and Maturity Profile)**, a practical framework for profiling AI systems. Instead of treating “AI” as a single, vague category, STAMP classifies implementations by **Role** (what the system does), **Agency** (how independently it operates), and **Environment** (where it runs and what is at stake). These dimensions generate a concise, one-page profile with a **Maturity Index** that summarizes risk, stage, evidence requirements, support level, and top risks—providing a clear, portable record that links system description directly to governance obligations.

Section 1 | Introduction

Artificial intelligence, or the ever-present shorthand, ‘AI’ now covers too much ground for one label to carry. The same shorthand describes predictive text that finishes an email, diagnostic systems that reshape medicine, and targeting software that guides autonomous weapons. When one word stretches to mean both conveniences and existential risks, it loses clarity. That semantic overload clouds debate, blurs accountability, and leaves stakeholders unclear about what these systems are and are not.

The stakes of ambiguity are more than linguistic. Opaque language fuels distorted expectations. Automation reshapes labor markets. Recommendation engines can amplify bias. Misinformation tools can tilt opinion. Advanced generative models blur the line between authentic and fabricated content. At the same time, analysts warn that capability can outpace governance, while hype cycles mix credible risks with Hollywood dystopias. These conflicting narratives make “AI” feel mythical rather than material, compressing decades of progress into a single buzzword.

When the term artificial intelligence was coined in 1956, it referred to techniques for making machines perform tasks that, if done by humans, would require intelligence[1]. Today, the label often obscures more than it explains. What we call AI is not an organism acquiring instincts of its own. It is a collection of engineered systems: algorithms encoded by people, trained on human data, and run on industrial infrastructure.

These systems do not have free will, instincts, or self-interest. When they appear to pursue objectives, it is because designers and data defined those objectives. Everyday verbs like “decide” or “choose” are metaphors of convenience, not evidence of intention. Fluent or creative outputs invite us to impute intelligence, but what we encounter is a statistical echo of our own humanity, recombined at scale.

This distinction matters for governance. When we treat AI as an autonomous “it” rather than an extension of human choices, we obscure accountability. Responsibility drifts from the people who set objectives, define data, and control deployment to the illusion of machine will. To prevent this drift, we need language that surfaces the real contours of a system: what risks it creates, what governance it requires, and the context in which it operates.

As practitioners, creators, policy makers, and members of society, we need to understand what we're actually building and deploying. We need clear models that work from system conception through its entire operational lifecycle. We need to grasp the complete profile of each implementation and use precise language to assess tradeoffs between capabilities and consequences. Most importantly, we need unambiguous chains of accountability for when outcomes diverge from intentions.

Scope of this paper

This paper presents a working methodology for profiling domain-specific AI implementations: deployed systems built to perform a defined function within a particular context. These implementations have clear boundaries, accountable owners, identifiable users, and measurable consequences. The framework is designed for focused applications rather than general-purpose platforms or multi-functional models like large language models that serve multiple roles without a primary function.

The methodology outlined in this paper is offered as a starting point for community refinement. The goal is to spark dialogue among practitioners about how we classify, govern, and take responsibility for the systems we're building. The framework, worksheets, and examples are presented openly, with the expectation that implementation will reveal gaps and improvements that can strengthen the approach for everyone.

What this paper proposes

This paper introduces a taxonomy model and profiling method to enable clear discussion and governance of AI systems. The taxonomy classifies systems by Role (function), Agency (operational independence), and Environment (deployment context and stakes). The System Taxonomy and Maturity Profile (STAMP) consolidates these dimensions into a one-page profile featuring a compact header that indicates system maturity, risk tier, required evidence, support level, and primary risks. This approach is offered as standardized artifact that bridges technical description and governance requirements across engineering, design, policy, audit, and leadership functions.

What this paper delivers

1. A three-part taxonomy: Role, Agency, Environment.
2. A repeatable protocol for producing a one-page profile that multiple stakeholders can read and come away with a shared understanding.
3. A maturity header that compresses stage, risk, evidence required, support level, and top risks.
4. Worksheets and tables you can paste into your process.

The underlying rationale is covered in Section 2, and Section 3 discusses the limitations of current categorical labels for operational governance. The methodology and framework are explained in Sections 4 and 5.

Section 2 | What Is This AI You Speak Of?

Across industries, organizations are deploying predictive, generative, and decision-making systems under one expansive label: "AI." This semantic overload creates a governance problem. The same term describes autocorrect and autonomous weapons systems. When one label covers both conveniences and existential risks, we cannot distinguish beneficial applications from genuine threats or embrace useful capabilities while governing dangerous ones.

The rush to integrate AI has created an environment where the label carries more weight than the utility. Organizations chase "AI initiatives" to satisfy boardroom demands and competitive pressures, often without shared expectations. Marketing motivations merge with executive mandates, leading teams to prioritize AI as a True North rather than build purpose-driven systems. This cycle rewards hype over substance, obscuring the real capabilities and limitations that matter for responsible deployment.

The reality beneath the mystique is simpler and more actionable. At their core, AI systems are engineered artifacts: algorithms designed by people, trained on human data, deployed to serve human objectives. They do not possess free will or self-interest. When they appear to act independently, those behaviors trace directly to designers' choices about data, objectives, and deployment parameters.

2.1 Constructed, Not Conscious

AI systems are built from three engineered components:

1. Models and algorithms that transform inputs and produce outputs.
2. Infrastructure that provides computational power.
3. Data that ultimately traces to human knowledge and decisions about what to measure, collect, and prioritize.

These elements have been combined with unprecedented computational power and scale to create remarkable capabilities that can appear almost magical. However, consciousness and independent will are not emergent properties of algorithmic complexity. The sophistication and fluency of outputs creates the illusion of understanding through statistical pattern recognition operating at industrial speed.

As Jaron Lanier warns, the real danger is not that machines will "wake up," but that humans will treat outputs shaped by our own choices as alien intentions. By anthropomorphizing systems, we excuse ourselves from accountability[3].

2.2 The Hall of Mirrors

When we encounter AI systems that seem to think, choose, or even deceive, we are not witnessing the emergence of digital consciousness. We are seeing ourselves reflected back through computational mirrors. Each apparent sign of independent thought has a mechanical explanation rooted in human design choices.

AI systems feel lifelike because they reflect and amplify human influence. Like mirrors, they transform what we place in front of them: our data, objectives, and biases. Understanding these patterns is crucial because they reveal the human fingerprints on every AI output, even when systems appear to act independently. What looks like evidence of a mind behind the glass is actually the predictable result of how we constructed the mirror itself.

> **Distortion:** Systems distort our inputs while appearing authoritative, like funhouse mirrors that bend reality. Sparse data gets smoothed into confident projections, incomplete information becomes false certainty. What appears to be editorial judgment is our own gaps reflected back as artificial confidence[4].
>
> **Amplification.** Small human choices become outsized effects through computational amplification. Minor weightings we embedded cascade through scale until they dominate outcomes. What looks like systemic bias is actually our subtle preferences magnified beyond recognition by scale.
>
> **Hallucination:** When there is nothing real to reflect, systems generate plausible fictions rather than acknowledge gaps. Pattern recombination fills voids where our training data was incomplete. What appears to be creative invention is statistical projection masquerading as knowledge[5].
>
> **Refraction.** The same system shows different faces depending on the angle we view it from. What looks like evolving personality is the same mirror reflecting different human objectives we defined for different contexts.

It is easy to mistake algorithmic aberrations for an independent mind when they reflect our own complexity back at us in unexpected ways. Understanding these distortion mechanisms returns accountability to where it belongs: with the humans who constructed the systems behind the glass, even as we struggle to recognize ourselves in what we've created.

If these behaviors trace to human choices about function, independence, and deployment context, then our governance frameworks should reflect that reality. Stakeholders need shared descriptors that answer basic questions: What does the system do? How independently does it operate? Where does it run and what are the stakes?

Existing classification schemes often obscure rather than clarify these distinctions. The next section examines where popular frameworks fall short and why we need more operational approaches to AI taxonomy.

Section 3 | The Need for Operational Classification

Why Classification Matters

How we classify AI systems shapes everything that follows: resource allocation, regulatory frameworks, public understanding, and accountability structures. A shared taxonomy enables executives, engineers, policymakers, and ethicists alike to reason together about the same systems. Without it, hype thrives, accountability blurs, and distinctions that matter for governance disappear.

The Limits of Current Approaches

The ANI/AGI/ASI Framework

The most common framework divides Artificial Intelligence capabilities into three broad evolutionary classifications:

- Narrow (ANI), General (AGI), and Superintelligence (ASI). ANI covers virtually all current systems, from spam filters to large language models[6].
- AGI represents proposed human-level capability that could learn flexibly across domains[6].
- ASI describes speculative systems that would far surpass human intelligence[6].

This framework emerged from research discourse and offers a compelling narrative of technological progression. However, it creates several problems for practitioners and policymakers.

1. It's entirely future-oriented, centering discussions around imagined capabilities rather than distinguishing today's diverse systems. Autocorrect and autonomous weapons both fall under "ANI" despite radically different governance needs.
2. The evolutionary framing implicitly suggests that systems naturally develop consciousness or desires as they grow more complex, deflecting attention from the human choices that shape outcomes.
3. It provides no actionable guidance for oversight, testing, or accountability.

Institutional Efforts

Recognizing these limitations, major institutions have developed more operational approaches.

- The **EU AI Act** creates risk-based categories with obligations scaling from minimal to prohibited uses, anchoring governance in context and consequences[7].
- **NIST**'s AI Risk Management Framework provides voluntary organizational processes around governance, mapping, measurement, and management[8].
- The **OECD** offers technology-neutral definitions emphasizing objectives, inference, and autonomy[9].
- Standards bodies like **ISO** develop shared vocabularies and lifecycle processes that facilitate technical interoperability, though these focus more on component definitions than functional distinctions[10].

Each effort advances the field within its domain. The EU approach effectively links risk to obligation. NIST strengthens organizational practice. International standards facilitate cooperation. Domain-specific frameworks, like automotive automation levels, create precise operational definitions that everyone in the sector understands[11].

Despite valuable contributions from regulatory frameworks, technical standards, and domain-specific approaches, practitioners still lack operational language and functional models that work across diverse stakeholder groups. The question isn't whether existing efforts are good or bad, it's whether teams can easily answer basic questions: What does this system do? How independently does it operate? Where does it run and what are the stakes?

These three dimensions - function, independence, and environment - surface the human choices that shape AI behavior while providing concrete anchors for governance decisions. The next section introduces a framework built around these dimensions.

Section 4 | STAMP Methodology (Taxonomy)

Much of today's discussion of "AI" treats it as if it were a single thing, a one-dimensional label applied to systems that behave very differently. To move beyond that, we need a way of describing systems with more depth and fidelity — one that makes their purpose, independence, and operating environment clearer, and improves the ability of technical, policy, and governance communities to work together.

The framework introduced here offers that three-dimensional approach. It defines three classification categories that together capture the defining qualities of a system:

- **Role:** what the system does.
- **Agency:** how independently it operates.
- **Environment**: where it runs and the boundaries that shape its exposure.

Together, these categories form the basis of the **System Taxonomy and Maturity Profile (STAMP)**, a method that links description directly to governance. By following a structured process, teams move from vague labels to a durable record that ties what a system does to the obligations and safeguards that apply.

Each category is paired with linked attributes in the appendix tables: **Definition, Diagnostic Markers, Risks, Governance Obligations, Default Safeguard**. Selecting a category and recording its attributes ensures classification surfaces what to monitor, the responsibilities to uphold, and the baseline safeguards to enforce. These attributes are not exhaustive prescriptions; they are reference points to support consistent practice. Different systems may present unique risks, obligations, or safeguards, and teams should document any additions or deviations alongside the defaults.

How to use this section

1. Complete 4.1 Role, 4.2 Agency, and 4.3 Environment using each step's How to choose rules.
2. After each choice, copy the linked attributes from the corresponding appendix table into the worksheet fields:
    - Role: fill 2.1–2.5 using Table A.1
    - Agency: fill 3.1–3.5 using Table A.2
    - Environment: fill 4.1–4.5 using Table A.3
3. Document a short rationale (3–5 sentences each) for Role and for Agency, noting observed behaviors, evidence of markers, consequence path, and why alternatives were ruled out.
4. Record evidence pointers (e.g., audit logs, approval records, test artifacts) as you go.

Section 5 converts these selections into a Maturity Index by lookup. No new analysis is required.

4.1 Role — What is the system's primary function?

Why this matters

Role defines the system's identity: what kind of function it performs in the world. It creates a common reference point for describing the system, aligning design choices, governance obligations, and communication across teams.

How to choose

Identify the first Role that accurately describes the system, working down the list in order. If more than one seems to apply, select the Role with the most immediate effect on outcomes — such as granting or denying access, changing a record, or initiating an action (grant/deny/change/actuate is the tie-breaker).

Question:

Identify your system's role by answering these questions in sequence. Stop at the first 'yes'

1. **Agentive** — Does the system initiate or execute actions that change digital or physical states?
2. **Evaluative** — Does it assign scores, labels, or ranks that grant/deny access, trigger thresholds, or carry consequences?
3. **Recommender** — Does it rank or filter items so users see some things first or most?
4. **Generative** — Does it produce new content (text, images, audio, video, code) beyond retrieval or labeling?
5. **Predictive** — Does it output probabilities or forecasts about future or latent states without imposing a label or threshold?
6. **Collaborative** — Does it co-author with a human in real time, adapting to feedback?
7. **Assistive** — Does it only suggest or guide while the human remains the sole decision maker, and none of the above apply?

Record in Worksheet

- 2.1 Role Selected
- 2.2 Diagnostic Markers (from Table A.1)
- 2.3 Risks (from Table A.1)
- 2.4 Governance Obligations (from Table A.1)
- 2.5 Default Safeguard (from Table A.1)

Once recorded, proceed to Agency.

4.2 Agency — How independently does the system operate?

Why this matters

Agency defines how independently the system operates when its outputs are applied. It shows whether results depend on human approval, can be overridden in real time, or proceed without feasible intervention. A clear Agency classification creates shared understanding of the system's level of autonomy, supporting design decisions, staffing, and governance.

How to choose

- Pick one level that matches how the system behaves at the point where the result actually occurs (e.g., access granted, a record changed, a device moved).
- If you are between two levels, choose the more independent one, since it carries stricter obligations in Table A.2.
- If behavior differs by environment or release, create a separate profile for each.

Question

How independently does the system operate?

- Advisory — Only provides guidance; human takes the action.
- Human-in-loop — Every action requires explicit human approval.
- Human-on-loop — System acts, but a human monitors and can intervene.
- Delegated — System acts automatically within predefined limits.
- Autonomous — System acts without feasible real-time human intervention.

Record in Worksheet

- 3.1 Agency Selected
- 3.2 Diagnostic Markers (from Table A.2)
- 3.3 Risks (from Table A.2)
- 3.4 Governance Obligations (from Table A.2)
- 3.5 Default Safeguard (from Table A.2)

Evidence tests (examples of verifiable artifacts)

- Human-in-loop: every action has an approval event ID linked to the execution ID.
- Human-on-loop: override mechanism exists with a defined service-level objective (SLO) for activation.
- Delegated: the system enforces a boundary ruleset; the latest conformance audit or exception log is available.
- Autonomous: documentation shows why real-time human intervention is infeasible (timing, physics, or scale), and monitoring confirms.

Once recorded, proceed to Environment.

4.3 Environment — Where does the system run and what is at stake?

Why this matters

Environment situates the system in its real-world domain and clarifies the consequences of failure. It sets the stakes and determines which governance obligations take priority.

How to choose

- Pick the Environment that best reflects where the system operates and the consequences of failure.
- If multiple environments apply, adopt the strictest resulting Risk and Evidence.
- If behavior differs across environments, create a separate profile for each.
- If the system operates across jurisdictions or as a platform layer, mark the Cross-jurisdictional / Platform flag and record the impacted areas.

Question

Where does the system run, and what is at stake if it fails?

- Consumer / Reputational — Affects user experience or brand trust.
- Financial / Economic — Affects money, credit, or access to opportunity.
- Civic / Legal — Affects rights, elections, or due process.
- Health — Affects patient care, outcomes, or medical data.
- Safety-critical — Affects human safety or poses risk of physical harm.

Record in Worksheet

- 4.1 Environment Selected
- 4.2 Diagnostic Markers (from Table A.3)
- 4.3 Risks (from Table A.3)
- 4.4 Governance Obligations (from Table A.3)
- 4.5 Default Safeguard (from Table A.3)

Once recorded, Section 4 is complete.

4.4 Outcome of Section 4

At the end of this section, the worksheet should contain:

- Role selected, with attributes copied from Table A.1.
- Agency selected, with attributes copied from Table A.2.
- Environment selected, with attributes copied from Table A.3.
- Rationale and evidence pointers for each classification.

This creates a structured identity record for the system.

Next step: Section 5 converts these selections into a five-part Maturity Index by lookup. No new analysis is required. These integrity rules are applied in Section 5 to compute **Risk** and **Evidence** from **Role**, **Agency**, and **Environment**, and to record the system's Maturity Index and derived Support Level.

Section 5 | From Taxonomy to Maturity

Section 4 established the three dimensions of a system profile: Role, Agency, and Environment. Those choices created a structured description of what the system does, how independently it operates, and the context in which it runs.

Section 5 takes those results and adds one more input — the system's Stage of deployment — to generate a compact header called the Maturity Index.

The Maturity Index summarizes a system in five plain fields:

- Risk Level — the governance level appropriate to the system's combination of Environment and Agency.
- Stage — where the system is in its lifecycle (Prototype, Pilot, Limited, Production, Critical).
- Evidence — the level of evidence expected for oversight.
- Support Level — how embedded and supported the system is inside the organization.
- Top Risks — the two most significant risks associated with the system.

Together, these form the header of the one-page profile. The goal is a consistent, scannable signal that can travel across engineering, product, audit, and policy without translation.

Steps

- Pick Stage.
- Use Table 5.1 to copy Risk Level and Evidence (after the short rule below).
- Use Table 5.2 to copy Support Level from Stage.
- Pick two Top Risks (one from Role, one from Environment).
- Apply the Integrity Rules (raise, never lower). Done.

Why this matters
Stage captures how far along the system is in deployment. A prototype in a lab and a critical production system are not governed the same way, even if they share the same Role, Agency, and Environment.

How to choose

Stage	**Definition**
Prototype	Internal experiment; test or synthetic data; no real users.
Pilot	Limited real users; time-boxed trial with rollback plan.
Limited	Restricted scope or audience; guardrails and on-call support in place.
Production	Broadly available in intended environment; SLOs and incident processes active.
Critical	Failure carries severe organizational or safety consequences; executive or regulator sign-off required.

Record in Worksheet
Enter the Stage in field **5.0**. In the Index, it prints simply as **Stage**.

Step 1 — Lookup Risk Level and Evidence

Why this matters
Risk Level and Evidence are set by where the system runs (**Environment**) and how independently it acts (**Agency**). Publishing the small rule below makes the lookup consistent and transparent.

Computation rule (publish before using the table)

- Set the **Environment risk floor**.
- Set the **Agency evidence floor** and any **Agency risk bump**.
- Apply the **Integrity Rules** (Step 4).
- **Risk Level = max(Environment floor, Agency bump, any Role-based floors).**
- **Evidence = max (Environment evidence floor, Agency evidence floor).**
  Values in Table 5.1 are floors. Integrity Rules may raise them; they never lower them.

How to use
Find the row that matches your Environment and the column that matches your Agency. Copy both values into the worksheet.

Table 5.1 — Environment × Agency Lookup (floors)

Environment	**Agency**			
	Human-in-loop	**Human-on-loop**	**Delegated**	**Autonomous**
Consumer / Reputational	Risk: Low • Evidence: Observed	Low • Documented	Trusted • Documented	Trusted • Audited
Financial / Economic	Trusted • Documented	Trusted • Audited	Trusted • Audited	Critical • Audited
Civic / Legal	Trusted • Documented	Trusted • Audited	Critical • Audited	Critical • Audited
Health	Trusted • Documented	Trusted • Audited	Critical • Audited	Critical • Audited
Safety-critical	Critical • Audited	Critical • Audited	Critical • Audited	Critical • Audited

Record in Worksheet

- **Risk Level → 5.1**
- **Evidence → 5.2**
- **Derivation notes (1 line) → 5.2a** *(which floor/bump/rule affected the result)*

- **Evidence pointer (URL/path/doc) → 5.2b*

Step 2 — Derive Support Level

Why this matters
Support Level shows how well the system is embedded and supported in the organization, which drives the maturity of its governance.

How to use
Use the Stage you selected in Step 0 to determine Support Level.

Table 5.2 — Stage → Support Level Map

Stage	Support Level	Example signals
Prototype	Experimental	Lab-only; synthetic data; no pager duty
Pilot	Supported	Named owner; rollback plan; basic monitoring
Limited	Supported	On-call rotation; guardrails enforced
Production	Operationalized	SLOs; incident playbooks; monitoring
Critical	Institutionalized	Executive or regulatory sign-offs; independent assurance

Record in Worksheet

- **Support Level → 5.3**

Step 3 — Select Top Risks

Why this matters
Every system has risks, but a concise signal requires focusing on the most significant. Pulling them from the Role and Environment tables avoids reinventing the list and keeps language consistent.

How to use
Review the risks listed under your chosen Role (Table A.1) and Environment (Table A.3). Select the two that most accurately represent your system's dominant risks.

Record in Worksheet

- **Top Risks → 5.4**

Step 4 — Apply Integrity Rules (raise, never lower)

To maintain consistency, apply these reminders before finalizing the Index:

- If you are **between two Agency levels** → pick the **more independent**.
- If the system **spans multiple Environments or jurisdictions** → keep the **highest Risk and highest Evidence**.
- If a **Secondary Role is Agentive** (grants/denies/actuates/enforces) → raise **Risk** to at least **Trusted** and **Evidence** to at least **Documented**.
- If a **feature flag or failover** can increase Agency (e.g., delegated → autonomous), profile the **highest Agency** reachable and note it in **5.2a**.

Record any notes on these adjustments in **5.2a Derivation notes**.

5.5 Outcome of Section 5

At this point, the worksheet should display the **Maturity Index** header at the top of the profile:

Maturity Index (Header)

- **Risk Level: ___**
- **Stage: ___**
- **Evidence: ___**
- **Support Level: ___**
- **Top Risks: ___, ___**

Beneath the header, the worksheet contains the taxonomy selections from Section 4, the copied attributes from the appendix tables, and any rationale and evidence pointers.

The result is a one-page system profile that is scannable, consistent, and durable across the system lifecycle. It replaces vague shorthand with a structured identity record that can travel across engineering, audit, and policy settings.

(Optional badge for audit readiness: Taxonomy integrity — Dual-rater [ ] • Rationale recorded [ ] • Evidence linked [ ])

Tiny terminology alignment (optional, zero logic change)

Appendix A — Detailed Taxonomy References

A.1 Roles (What the system does)

Role	Definition	Diagnostic Markers	Risks	Governance Obligations	Default Safeguard
Agentive	Acts directly to change digital/physical states without per-action approval.	Initiates processes independently; changes states in software/physical world; harms manifest as actions.	Unsafe actions; goal misalignment; fast cascading harms.	Pre-deployment validation; redundancy and fail-safes; real-time monitoring.	Validation + live monitoring with emergency override.
Evaluative	Assigns scores, ranks, or classifications with consequences.	Produces scores or categories; impacts access to resources/rights; includes/excludes based on threshold.	Bias; exclusion; opacity.	Fairness reviews; explainability for users; appeals and redress.	Individual explanations with accessible appeal path.
Recommender	Filters/ranks information to shape attention and exposure, often implicitly.	Determines what is seen first or most often; filters/ranks content; primary impact is attention shaping.	Echo chambers; harmful amplification; hidden influence.	Transparent ranking criteria; guardrails on amplification; user controls.	Reveal ranking criteria and provide effective opt-outs.
Generative	Produces new content by recombining patterns from data; not retrieval or labeling.	Produces text, images, audio, video, or code; outputs are recombinations; removing system removes content capacity.	Hallucination; IP/copyright risk; misinformation or impersonation.	Provenance and labeling; review for high-stakes contexts; misuse controls.	Label AI-generated outputs and provide provenance metadata.
Predictive	Estimates future states or probabilities.	Outputs forecasts/probabilities; time/scenario-oriented; downstream reliance.	Miscalibration; dataset shift; cascading errors.	Regular recalibration; stress testing; oversight in high-stakes contexts.	Scheduled recalibration and scenario stress tests.
Collaborative	Works interactively with a person during creation or analysis; extends cognition or creativity.	Adapts in real time to feedback; shares authorship of intermediate outputs; human finalizes the result.	Blurred authorship; accountability deflection; skill atrophy.	Attribution and authorship standards; clear division of labor; audit trails.	Attribution and audit trail enabled by default.
Assistive	Surfaces suggestions or guidance without taking binding action. The human remains the decision maker.	Suggests rather than enforces; human can complete task without it; value is efficiency, error reduction, accessibility.	Over-reliance; false sense of accuracy; unequal accessibility.	Transparency of limits; usability testing with diverse users; always provide override.	Human override and error-rate disclosure.

A.2 Agency (How independently the system operates)

Agency Level	Definition	Diagnostic Markers	Risks	Governance Obligations	Default Safeguard
Advisory	Provides info or recommendations only; never acts.	Outputs are advice, not actions; human execution required; harm from misunderstanding, not system action.	Misinterpretation; over-reliance; drift from intended use.	Confidence reporting; clear decision rights; user training.	Confidence reporting with documented assumptions.
Human-in-loop	Proposes actions but requires explicit approval.	Cannot execute without approval; traceable checkpoints; oversight on approvals.	Rubber-stamping; workflow bottlenecks; unclear accountability.	Formal approval mechanisms; immutable logs; reviewer training.	Approvals with immutable logs and clear criteria.
Human-on-loop	Acts without pre-approval but under continuous monitoring.	Operates autonomously between interventions; override possible; risk rises if supervision lapses.	Inattentive supervision; ambiguous override thresholds; alarm fatigue.	Real-time monitoring; clear protocols; operator staffing/training.	Real-time monitoring with override thresholds.
Delegated	Acts within predefined bounds, oversight is periodic.	Executes automatically within limits; oversight exception-based; safety depends on boundaries.	Boundary failure; specification gaming; slow error detection.	Guardrails and rate limits; boundary audits; certification for expansion.	Guardrails with boundary-compliance audits.
Autonomous	Operates beyond narrow bounds where oversight is infeasible.	No meaningful oversight in real time; not tightly constrained; humans cannot intervene fast enough.	Unsafe/irreversible actions; emergent behavior; cascading harms.	Prohibition or extreme caution; maximum validation; explicit liability frameworks.	Deploy only with maximum validation and liability controls; prohibit where unmitigable.

A.3 Environment (Where it runs and what is at stake)

Scope	Definition	Diagnostic Markers	Risks	Governance Obligations	Default Safeguard
Closed / Isolated	Runs without external connectivity (air-gapped, offline, embedded).	No network interfaces; limited to preloaded data; operates inside secure enclave.	Blind spots from outdated models; weak monitoring; insider abuse.	Periodic offline validation; access controls; manual update approvals.	Independent audit of update/validation cycle.
Enterprise / Controlled	Runs within a secure organizational perimeter.	Accessible only to authenticated org users; hosted in private cloud or internal network.	Insider bias; uneven governance across departments; leakage if perimeter fails.	Data governance and access logs; org-wide policy alignment; incident response readiness.	Quarterly governance review + perimeter monitoring.
Public / Open	Accessible by general users over internet/app channels.	Public API, web, or consumer-facing app; unbounded user base.	Scale of harm; adversarial misuse; reputational damage.	Transparency requirements; red-teaming; public disclosure of limits.	Mandatory terms of use + abuse monitoring.
Device-embedded / Embodied	Runs on hardware that interacts with the physical world.	Integrated into IoT, robotics, vehicles, or medical devices.	Physical safety risks; unintended actuation; liability ambiguity.	Safety case documentation; certification; incident fail-safe.	Human override/kill switch.
Critical Infrastructure	Deployed where failures cascade across systems (energy, aviation, defense, utilities).	Designated as "critical" by regulators; high-dependency networks.	Catastrophic outages; systemic security threats; national security risk.	Regulator notification; resilience testing; continuity planning.	Independent safety regulator sign-off.
Cross-jurisdictional / Platform layer **(optional)**	Provides core services consumed by downstream systems, or spans multiple legal regimes.	API-based services with many external adopters; multinational deployment.	Liability diffusion; conflicting legal obligations; uneven compliance.	Cross-border compliance mapping; contractual governance with integrators.	Compliance registry and jurisdictional impact review.

Appendix A.4 — STAMP Profile Template / Worksheet

A) TAXONOMY INPUTS (from Section 4)

4.1 **Role:**

- ☐ Agentive
- ☐ Evaluative
- ☐ Recommender
- ☐ Generative
- ☐ Predictive
- ☐ Collaborative
- ☐ Assistive

4.2 **Agency:**

- ☐ Advisory
- ☐ Human-in-loop
- ☐ Human-on-loop
- ☐ Delegated
- ☐ Autonomous

4.3 **Environment:**

- ☐ Consumer / Reputational
- ☐ Financial / Economic
- ☐ Civic / Legal
- ☐ Health
- ☐ Safety-critical

B) COMPUTE RISK AND EVIDENCE (Step 1 rule + Table 5.1)

5.1 Risk: _____

5.2 Evidence: _____

5.2a Derivation notes (one line): _____

5.2b Evidence pointer (URL, path, or doc): _____

C) DEPLOYMENT MATURITY (Step 0 + Table 5.2)

5.0 Stage: ☐ Prototype ☐ Pilot ☐ Limited ☐ Production ☐ Critical

5.3 Support Level (from table): _____

D) TOP RISKS (Step 3) — exactly two

1) _____ (from Role)

2) _____ (from Environment)

E) OPTIONAL BADGE

Taxonomy Integrity: ☐ Dual-rater ☐ Rationale recorded ☐ Evidence linked

A.6 Process Integrity (taxonomy integrity) To support consistency and auditability:

- Dual-rater classification. Two independent raters classify Role and Agency. Environment may be single-rater if unambiguous.
- Disagreement ladder. If raters disagree, discuss once; if unresolved, escalate to a named steward who makes the final call. Record the decision, steward, and any Primary/Secondary Role change.
- Rationale capture. Provide a 3–5 sentence rationale for Role and Agency covering observed behaviors, evidence of markers, consequence path, and why alternatives were ruled out.
- Version control. Increment the worksheet version whenever Role, Agency, or Environment changes, or when a default safeguard is overridden. Maintain a short changelog.

References

[1] J. McCarthy, M. Minsky, N. Rochester, and C. E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1956.

[2] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Hoboken, NJ, USA: Pearson, 2020.

[3] J. Lanier, *You Are Not a Gadget*. New York, NY, USA: Knopf, 2010.

[4] Shah, D., Schwartz, R., & Hovy, D. (2020). Predictive biases in natural language processing models: A conceptual framework and overview. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5248-5264.

[5] Z. Ji, N. Yu, S. Xu, Y. Yang, H. Yu, and Y. Wu, "Survey of hallucination in natural language generation," *ACM Comput. Surveys*, vol. 55, no. 12, pp. 1–38, 2023.

[6] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford Univ. Press, 2014.

[7] European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on artificial intelligence (Artificial Intelligence Act)," *Official Journal of the European Union*, 2024.

[8] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD, USA: U.S. Dept. of Commerce, 2023.

[9] Organisation for Economic Co-operation and Development, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). Paris, France: OECD, 2019.

[10] ISO/IEC 22989:2022, *Information Technology — Artificial Intelligence — Concepts and Terminology*. Geneva, Switzerland: International Organization for Standardization, 2022.

[11] SAE International, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, SAE Standard J3016_202104, Apr. 2021.

How to Cite This Paper
Holcomb, Paul. *Don't Call Me Artificial: The STAMP Methodology for Creating Standardized System Profiles for AI Implementations.* 2025.