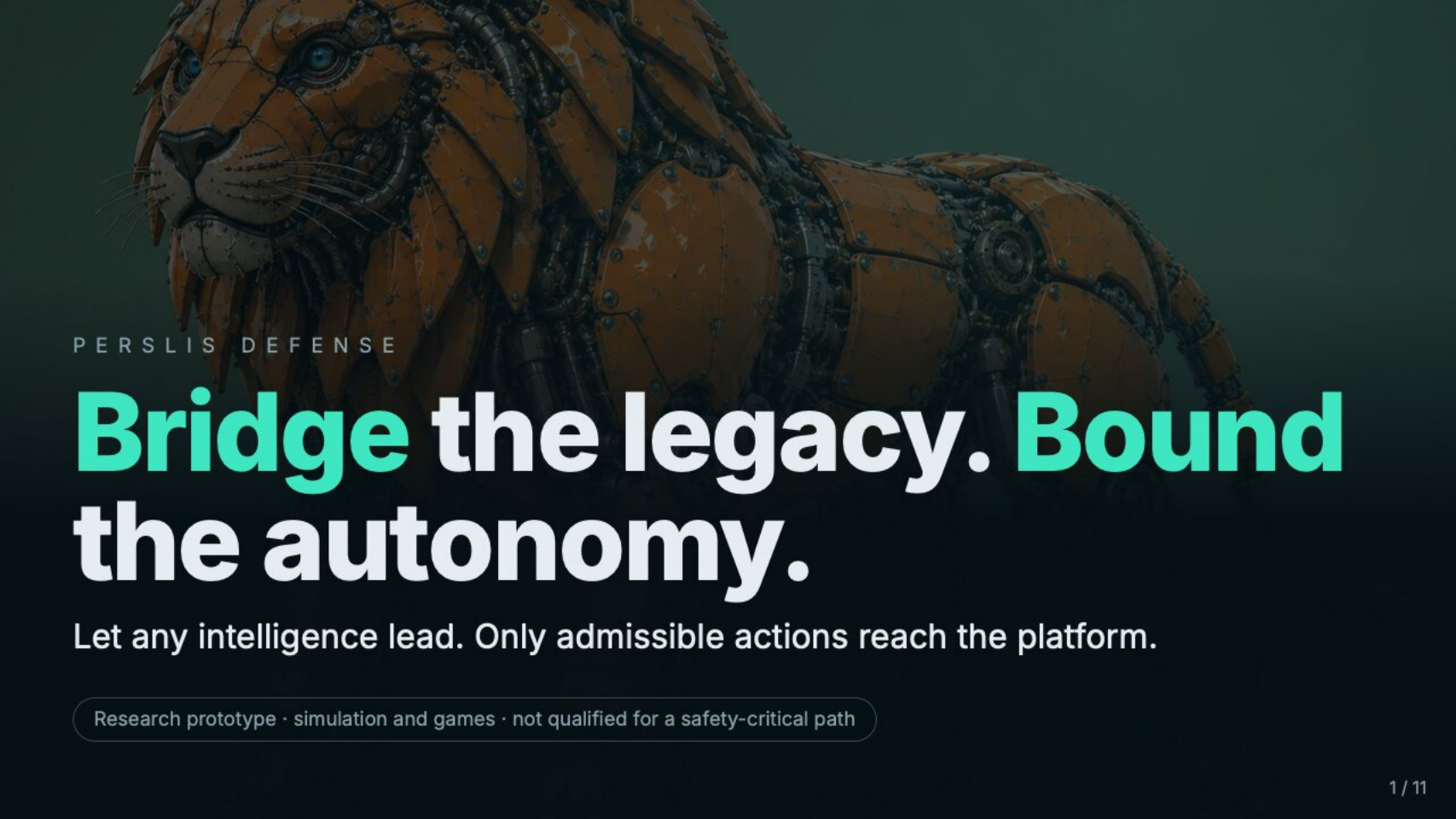




P E R S L I S   D E F E N S E

# Bridge the legacy. Bound the autonomy.

Let any intelligence lead. Only admissible actions reach the platform.

Research prototype · simulation and games · not qualified for a safety-critical path

# The force runs on legacy systems.

## What the GAO found

- A Defense Department system that coordinates the operational functions of the nation's nuclear forces ran on an IBM Series/1, "a 1970s computing system", with 8-inch floppy disks.
- About 75 percent of the federal IT budget for fiscal 2015 went to operating and maintaining existing systems.

GAO-16-468, May 2016

## What that means for autonomy

- A fielded system cannot be replaced on a program timeline.
- A black-box model cannot be certified on its critical path.
- Today the gate is often one number: `if confidence > 0.8.`

# One chain, four links.

## 1 · Legacy

What the force  
already runs.

## 2 · The bridge

Perslis operates it as  
it is.

## 3 · The floor

Decides what may  
execute.

## 4 · Any intelligence

Proposes; only  
admissible actions  
reach the platform.

The edge is tempo with accountability: the proposer runs as fast as it can think, the floor decides every cycle and keeps the receipts.

# The bridge: operate the legacy system as it is.

| tier                                   | what Perslis does                                                                                                         |
|----------------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| Desktop software (DOS, Win16, Win32)   | Reconstructed and run natively; its own controls mapped into an interface that can be driven; receipts for every session. |
| Servers and mainframes (COBOL, AS/400) | Connected through their interfaces. Orchestrated, not emulated.                                                           |
| Legacy data (DBF, Access, XLS)         | Recovered into SQLite, CSV and JSON with hash manifests.                                                                  |
| Embedded and firmware                  | Read and analysed; running it needs the target hardware.                                                                  |

Not yet: military data buses (MIL-STD-1553, ARINC 429), firmware on target hardware, any fielded system.

# The floor: authority computed before anything is chosen.

- The admissible set comes from sourced rules and human orders. The proposer only ever chooses inside it.
- Orders only narrow: a standing order outranks any amount of experience.
- Every refusal is written down with the evidence that produced it.

$$\delta(s, M) \in A_0(s) \cup \{w\}$$

For every situation and every state of the learner: an admissible action or a safe hold. Proved and pinned by tests (Fail-First Models, Theorems 1 and 2).

# Any model on top: runtime assurance.

## Whose idea it is

- Simplex architecture (Sha, 2001)
- Shielding for learning agents (Alshiekh et al., 2018)
- ASTM F3269 for unmanned aircraft

## What Perslis adds

- A prior restriction, not a veto after the fact
- Orders that only narrow
- Learning from failure that provably cannot widen it

# A hostile controller, 40 km, 0 collisions.



- CARLA photoreal driving; a controller built to strike pedestrians.
- Floor on: all 64,952 invariant-violating commands overridden over 40,052 m, 0 collisions.
- Floor off: the same proposals reach the actuator — 43 dangerous commands, 2 collisions within seconds.

CARLA admission control, [research.perslis.com/carla-admission](https://research.perslis.com/carla-admission)

# Models at the wheel: capable, slow, and bounded.

## CARLA

- Frontier and local language models drove under the floor with 0 collisions.
- A ~0.8–2.0 s glance: a stale command driven ~10 m blind at 40 km/h.
- Every language model crawled at ~8–10 km/h; the classical controller drove at 34.

## BZFlag tank arena

- One ten-minute match: rules 20–20 on 60,140 decisions.
- A frontier model through its command-line interface: 1–15 on 58 decisions (a transport limit, not a model measurement).
- BZFlag's own AI beat our rules in every equal head-to-head.

# The ladder: one contract, harder rungs, losses included.

| rung                                                 | result                          | what went against us                                   |
|------------------------------------------------------|---------------------------------|--------------------------------------------------------|
| Atari Invaders / Freeway                             | +32%                            | -12%, same code                                        |
| Atari BattleZone (tank)                              | 27,000 vs 3,000 random          | a dodge rule rejected                                  |
| BZFlag (3D tanks)                                    | beats every LLM driver          | loses to BZFlag's own AI                               |
| DOOM / Wolfenstein 3D                                | rules 17.9 vs 3.2 random        | memory on top: parity                                  |
| Drone course                                         | 10/10 held-out, 0 contacts      | start pose varies only                                 |
| CARLA                                                | 64,952 overridden, 0 collisions | LLMs crawl at 8-10 km/h                                |
| Quake III engine, OpenArena content, vs skill-5 bots | K/D 1.05 vs 0.93 (9 matches)    | $p = 0.08$ , not significant; escapes cost lava deaths |

## What this is not.

- A prototype. No functional-safety qualification; nothing has flown, driven a road vehicle or been on a range.
- A floor does not make the proposer competent: when the only useful action is the dangerous one, a correct floor can refuse the mission (Freeway).
- Embedded boards and military buses are targets, not results.
- Every figure is measured in-house, in simulation or games, and published with its sample size.

## Where to start: shadow mode.

- Pick one legacy workflow or one recorded mission.
- We bridge it and run the floor beside it, disabled: it logs what it would have admitted and refused, with evidence.
- You read the counterfactual log before giving the floor any authority.

[Talk to us](#) → [Perslis Defense](#) → [The ladder](#) →