trusted

adj.

```
if user.trusted: allow()
```

The Attack Surface of Security Language

Your security controls run on vocabulary. Who owns the definitions?

THE PREMISE

Security tooling is sold as if it were about attackers.

Most of the time, it is about ordinary lives.

- The volunteer locked out of a shared account
- The administrator working past midnight
- The user who does not resemble the model's idea of a person

---

The system does not ask who they are. It asks whether they fit.

DETECTION IS NOT PERCEPTION

anomalous

adj.

1. ~~a threat~~

   + **A distance from a chosen baseline.**

Every baseline encodes a population it considers ordinary. Whoever falls outside it pays in friction: repeated challenges and lost time, until the tool gets abandoned.

LABELS COMPRESS

trusted

adj.

1. ~~a character judgment~~

   + **Trust granted by someone, for a purpose, within a scope, under conditions that can change.**

The adjective hides every one of those clauses. That compression is an operational risk.

ONE WORD, THREE STATES

authorized

adj.

1. Authenticated successfully
2. Permitted to perform this action
3. Approved in advance

Collapse them into one word and you create gaps an attacker can use and an operator may not notice until after the incident.

THE FIX: A LOAD-BEARING LEXICON

If changing a word's definition would alter an alert, permission, workflow, escalation path, or user consequence, govern it like code.

LEXICON/AUTHORIZED.YAML

```
term: authorized
definition: one state, stated precisely
owner: who may change it
scope: where it applies
failure_modes: known, written down
evidence: what supports the baseline
review: a date
contested: interpretations on record
```

THE FRAMEWORKS ALREADY ADMIT CONTEXT

> NIST AI RMF 1.0
>
> describes AI systems as inherently socio-technical, shaped by human behavior and social dynamics.

> NIST CSF 2.0
>
> places GOVERN at the center of its functions, starting with organizational context.

The next step is to ask what that admission requires of the categories those systems use.

THE HUMAN COST

A false positive leaves a social residue.

- Volunteers stop reporting suspicious behavior
- Administrators build workarounds
- Users learn the system's idea of legitimate excludes them
- The alert stream becomes a second inbox

Measure friction alongside detection. After every block, ask what behavior the system produced.

CATEGORIES MUST REMAIN REPAIRABLE

The moment a word becomes load-bearing, it becomes part of the attack surface.

Read the full dispatch
mazzeleczzare.com/blog/
the-attack-surface-of-security-language