

we can't prove "DAIS-10 is always better" in an absolute sense, but we **can** show:

- Under realistic AV assumptions,
- For a clear safety objective,
- DAIS-10's **meaning-based rule** has **strictly lower expected harm** than a pure probability threshold rule.

I'll keep it tight and focused on the **calculation logic**.

1. Setup: a minimal safety model

We model a single critical scenario:

- A pedestrian may or may not be behind an occlusion (e.g., truck).
- The AV must choose: **Brake** or **Don't brake**.
- If a pedestrian is there and the car doesn't brake → high harm.
- If no pedestrian is there and the car brakes → minor cost (comfort / time).

Define:

- P: event "pedestrian exists behind obstacle"
- $p = \Pr_{f_0}(P | \text{sensor data})$: AV's estimated probability
- CFN: cost of a **false negative** (pedestrian exists, no brake) — very large
- CFP: cost of a **false positive** (no pedestrian, brake) — small

Assume $\text{CFN} \gg \text{CFP}$ (life vs inconvenience).

2. Baseline AV rule (probability threshold)

Typical AV decision logic:

- Choose a threshold $\theta \in (0, 1)$.
- If $p \geq \theta \rightarrow$ **Brake**
- If $p < \theta \rightarrow$ **Don't brake**

Expected cost under this rule:

$$E[\text{CAV}] = \Pr_{f_0}(P) \cdot E[C|P] + \Pr_{f_0}(\neg P) \cdot E[C|\neg P]$$

Conditioned on the posterior estimate p , the decision is:

- If **Brake**: expected cost $\approx \text{CFP}$ (even if P is true, braking is the safe action we want).
- If **Don't brake** and P is true: cost CFN.

So, given a particular p :

- If brake: CAV, brake \approx CFP
- If don't brake: CAV, no brake= $p \cdot$ CFN

The **probability-optimal** decision (minimizing expected cost) is:

- Brake if $CFP < p \cdot CFN \Rightarrow p > CFP/CFN$

So the “rational” threshold is:

$$\theta^* = CFP/CFN$$

If the AV uses a higher threshold (which many do for “comfort” / fewer interventions), it **accepts more risk than cost-optimal**.

3. DAIS-10 rule in the same scenario

DAIS-10 doesn't care only about p . It cares about **meaning** and **recent history**.

Let:

- S: DAIS-10 semantic score for “pedestrian presence” (0–100)
- Previously visible pedestrian \rightarrow Meaning-Defining \rightarrow S initially high (e.g., 95)
- Each frame where the pedestrian is not directly observed \rightarrow apply fading, but **not to zero**

Example fading model (DIFS-10):

- First missing frame: $S = 95 - 5 = 90$
- Second missing frame: $S = 90 - 10 = 80$
- Third missing frame: $S = 80 - 20 = 60$

Define a **semantic critical threshold**:

$$S_{crit} = 70$$

DAIS-10 decision rule:

- If attribute is Meaning-Defining and $S \geq S_{crit} \rightarrow$ **Treat as “still there” \rightarrow Brake / high-governance action.**
- Only when $S < S_{crit}$ may you downgrade governance.

So in the same occlusion case:

- AV probability: $0.42 \rightarrow 0.31 \rightarrow 0.12 \rightarrow 0.05$
- DAIS-10 meaning score: $95 \rightarrow 90 \rightarrow 80 \rightarrow 60$

On frame 2:

- AV might already be below its operational threshold and decide **not to brake**.
- DAIS-10 still has $S=80 \geq 70 \rightarrow$ **enforce braking / caution**.

4. Numerical comparison under reasonable parameters

Let's pick some realistic numbers:

- $CFN=1,000,000$ (catastrophic harm)
- $CFP=1,000$ (hard brake, discomfort, minor cost)

Then:

$$\theta^* = CFP/CFN = 1,000/1,000,000 = 0.001$$

A **strictly risk-minimizing** AV should brake if $p > 0.001$. In practice, many AV implementations effectively behave with a **much higher behavior threshold** (e.g., 0.1, 0.2, 0.3) to avoid “over-reacting.”

Let's compare behavior at $p=0.12$:

- Many AV stacks might decide: 0.12 is “low,” so **no brake**.
- DAIS-10 still treats the pedestrian as Meaning-Defining with high semantic score (e.g., 80) \rightarrow **Brake**.

Expected cost if AV chooses **no brake** at $p=0.12$:

$$E[CAV|no\ brake, p=0.12] = p \cdot CFN = 0.12 \cdot 1,000,000 = 120,000$$

Expected cost if DAIS-10 chooses **brake**:

$$E[CDAIS|brake] \approx CFP = 1,000$$

Comparison at that decision point:

$$120,000 \gg 1,000$$

So, under any situation where a system chooses “no brake” at p such that

$$p \cdot CFN > CFP$$

the DAIS-10 rule (which continues to enforce caution because of its high semantic score) has strictly lower expected harm.

This inequality is exactly the condition:

$$p > CFP/CFN$$

If an AV behaves as if its effective threshold $\theta_{AV} > \theta^*$, then a DAIS-10-style conservative semantic rule **dominates it in expected safety cost**.

5. Where DAIS-10 is mathematically “better”

We can say this clearly:

If:

1. The cost of a false negative is much larger than the cost of a false positive ($CFN \gg CFP$), **and**
2. The AV behaves as if its action threshold $\theta_{AV} > \theta^* = CFP/CFN$ (i.e., it waits for “high confidence” before braking), **and**
3. DAIS-10 keeps **Meaning-Defining** attributes in a “governed” state (high S) longer than the AV’s probability logic would,

then:

- At the decision points where AV chooses **no brake** and DAIS-10 chooses **brake**,
- The **expected safety cost of DAIS-10 is strictly lower**:

$$E[CDAIS] = CFP < p \cdot CFN = E[CAV]$$

That is a **formal inequality**, not hand-waving.

6. What’s actually being “proven”

We’re not proving “DAIS-10 is universally superior in all possible universes.”

We are showing:

- Given a standard risk-based cost model,
- Given realistic cost ratios in AV safety,
- Given current industry behavior (high implicit thresholds),
- And given DAIS-10’s conservative treatment of **Meaning-Defining** attributes,

then **DAIS-10’s semantic governance rule yields lower expected catastrophic harm** than a naive or comfort-tuned probability threshold.

That *is* a meaningful mathematical argument.