# Autonomous Trustworthy Agents (ATA)

06.11.2025
Engineering Kiosk Alps

Otera
David Peer
david.peer@otera.ai
**Machine Learning Researcher**

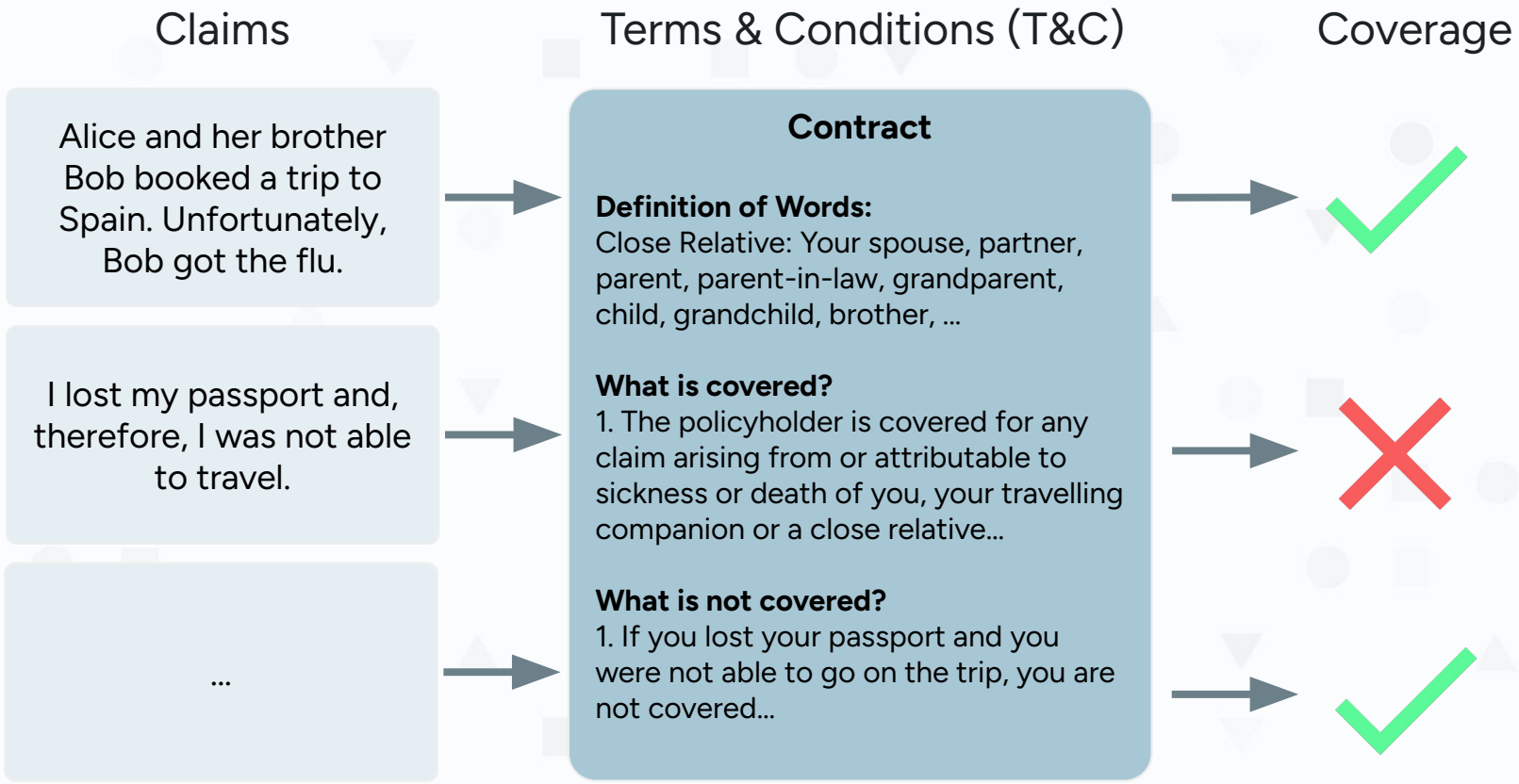
"... Zeit für 800 Folien"

# Otera

Previously DeepOpinion

Autonomy, Control, and Velocity for processes where quality, consistency, and trust are non-negotiable.

# Insurance - Claims Processing



# “Raw” GenAI for Claims Processing

# Claims Processing with “raw” GenAI

**Analysis of Coverage**

**Relative**

Mother, mother-in-law, father, father-in-law, sister, sisterin-law, brother, brother-in-law, wife, husband, civil partner, daughter, daughter-in-law, son, son-in-law, grandparent, grandparent-in-law, grandchild, step parent, step child, step sister, step brother, foster child, partner or fiancé/fiancée of an **insured person**.

the death, bodily injury or **illness** of:

llation due to the illness of a relative (cousin) and a
ged to travel (friend) is covered under Section 1 -
ation to suggest any major exclusion applies.

**yes**

1. **Cousin got the flu:** A cousin is explicitly defined as a **relative**. Illness of a relative is a covered reason for cancellation. The term "flu" falls under "illness" or "disease".
2. **Close relative and friend got sick:** This would cover the *relative* (cousin) and a *whom you are travelling* (friend).

T&C.pdf
Page 10

Enter a prompt for Gemini

Tools

Can we really trust[1] “raw” GenAI?

# Trustworthy[1] AI

- Accuracy & Stability
- Transparency & Explainability
- Human Agency & Oversight
- Privacy & Security
- Accountability
- Fairness & Non-Discrimination

# ~~Solution~~ -
# Generic Scaling Law for Everything

**More Data**

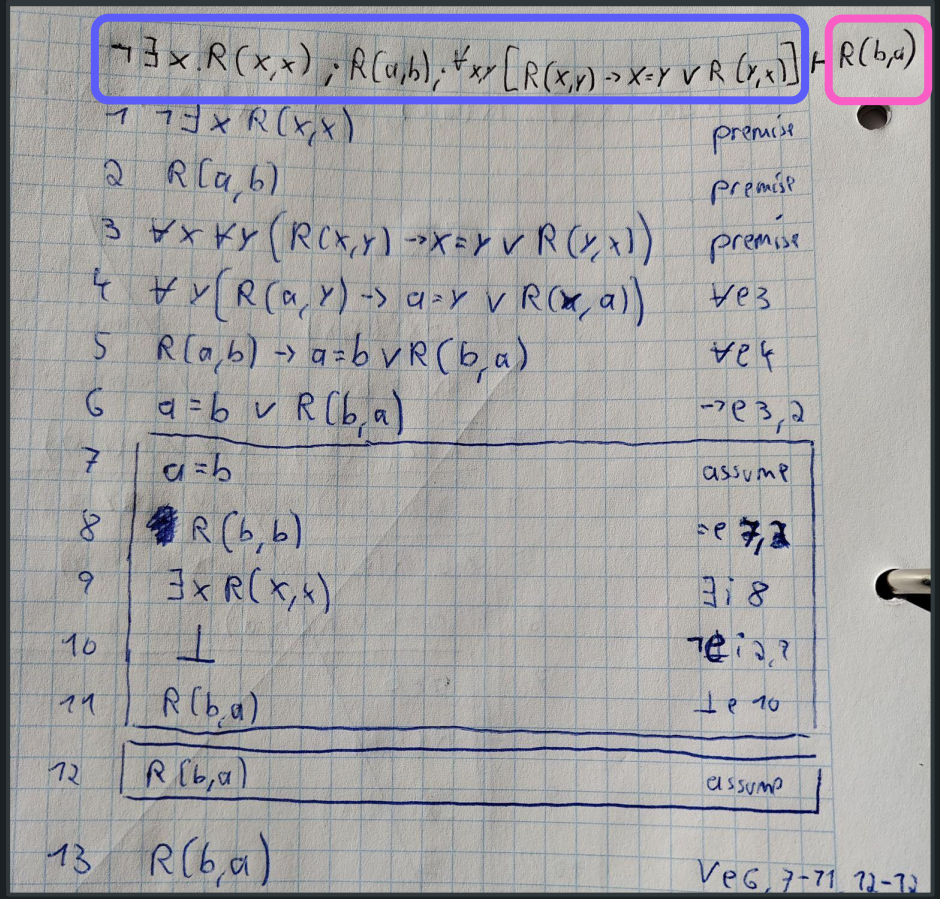

**More Compute**

# Autonomous Trustworthy Agents (ATA)

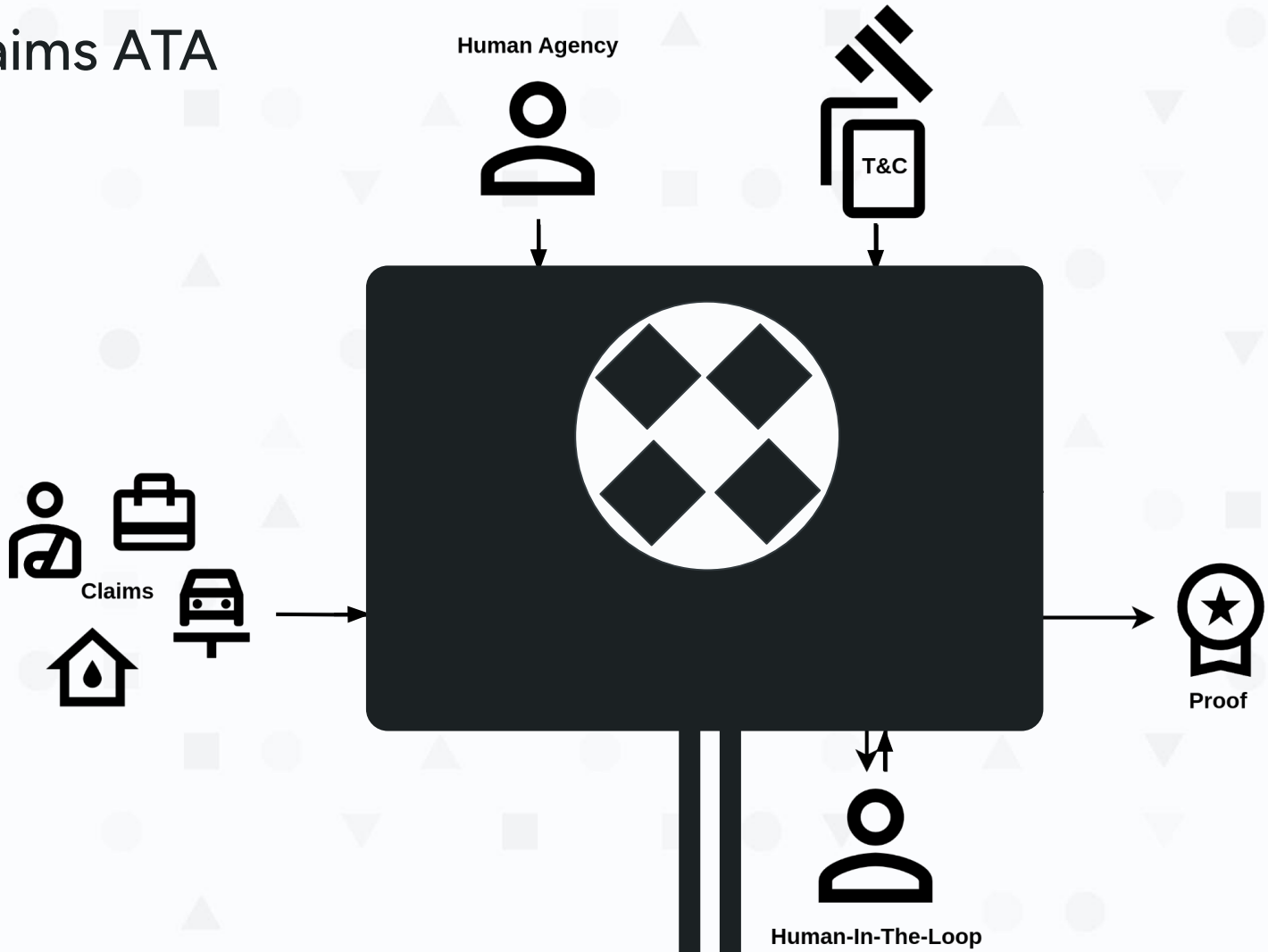
# Idea...

Maybe we can use Many-Sorted First-Order Logic to formalize T&Cs?

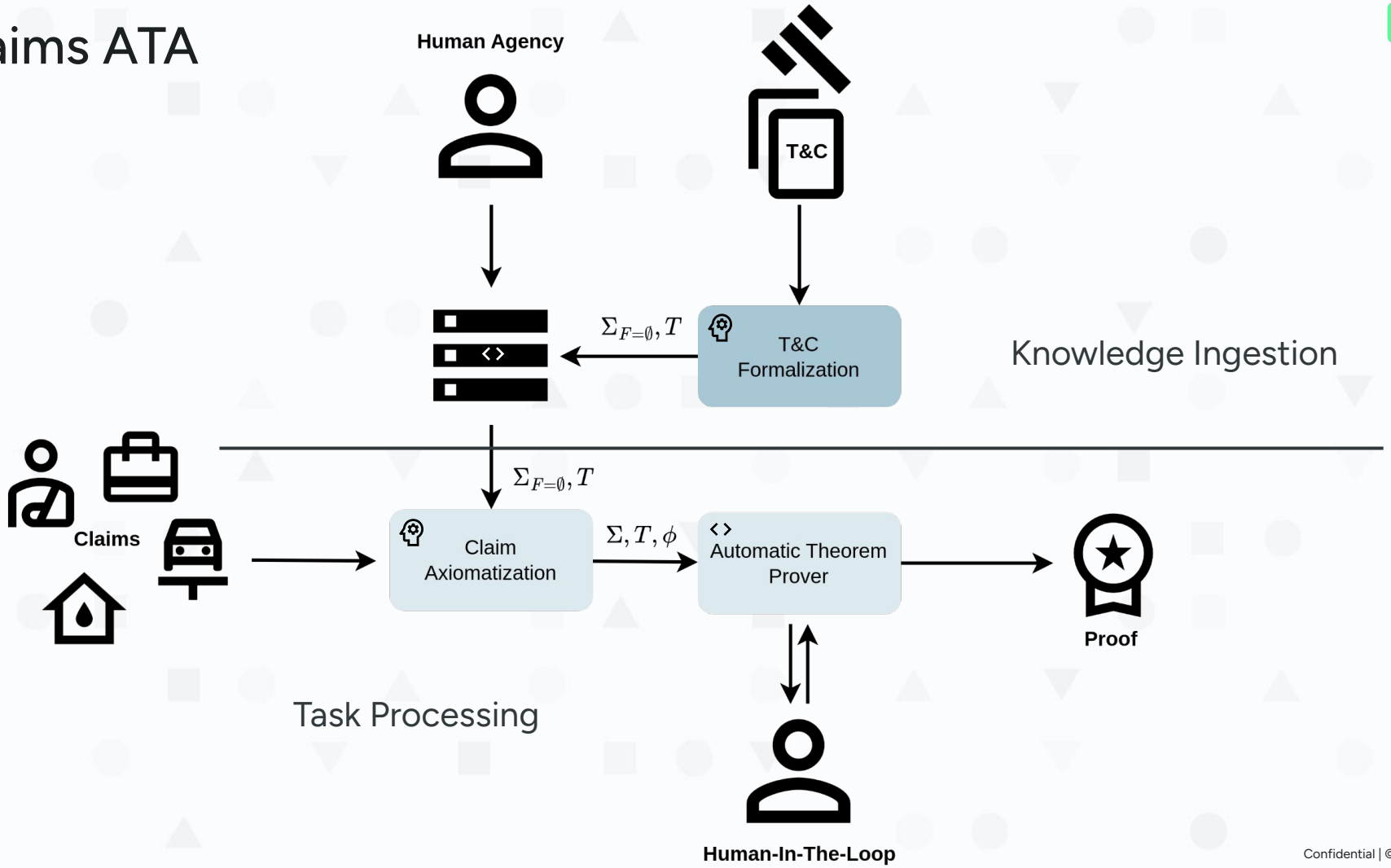
**Knowledge Base → T&C, Claim**
**Goal → Covered?**



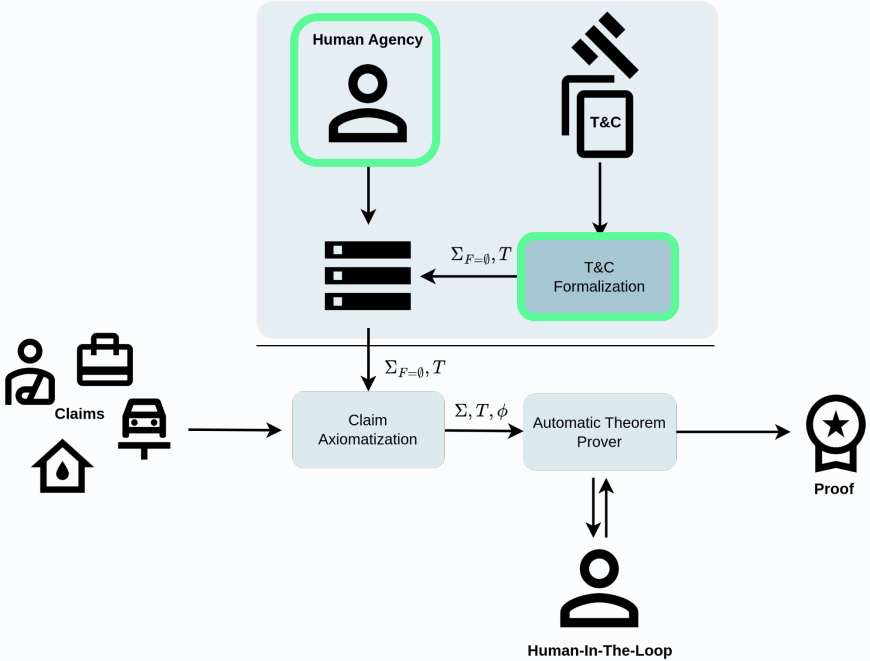
# Claims ATA



# Claims ATA



# Knowledge Ingestion




## Predicates & Sorts

`is_relative(p:Person, r:Person):`
Whenever person r is the spouse, partner, mother, sister, … of person p

…

---

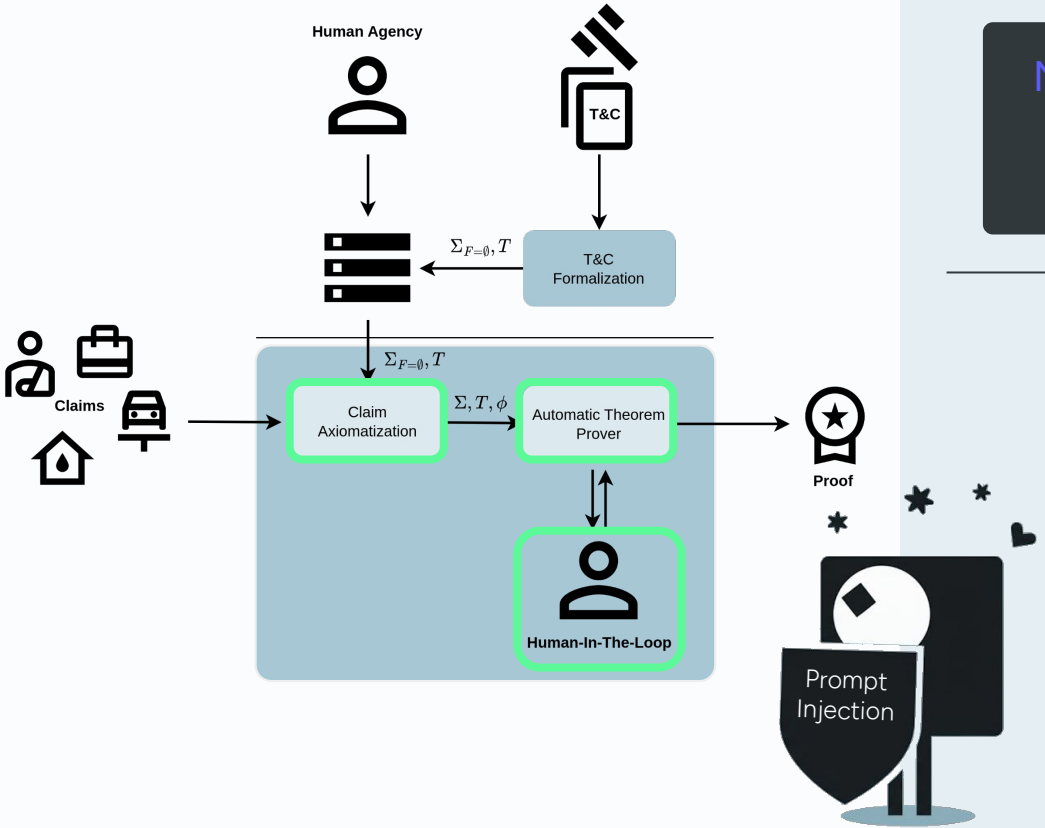
The policyholder is covered for any claim arising from or attributable to sickness or death of (1) you (2) your travelling companion (3) a close relative.

**?=**

**Formula**

$$\forall p, r.\ (\texttt{is\_sick(p)}) \vee (\texttt{is\_dead(p)}) \vee (\texttt{is\_sick(r)} \wedge \texttt{is\_travelling\_with(r, p)}) \vee (\texttt{is\_dead(r)} \wedge \texttt{is\_travelling\_with(r, p)}) \vee (\texttt{is\_sick(r)} \wedge \texttt{is\_relative(r, p)}) \vee (\texttt{is\_dead(r)} \wedge \texttt{is\_relative(r, p)}) \rightarrow \texttt{is\_covered(p)}$$

# Task Processing




Alice and her brother Bob booked a trip to Spain. Unfortunately, Bob got the flu...

Named Entity Recognition (NER)
Relation Extraction (RE)
is all you need

```
is_covered(ALICE)
```

Automatic Theorem Prover

Valid

Invalid

# Autonomy, Control, and Velocity for processes where quality, consistency, and trust are non-negotiable.

## Accuracy

(Zero Shot)

| | Base | ATA |
|---|---|---|
| Travel | 66% | **73%** |
| Electro. | 73% | **74%** |
| Dental | **92%** | 89% |

## Stability

(Standard Deviation)

| | Base | ATA |
|---|---|---|
| Int. | ±1.2 | **±0.0** |
| Ext. | ±0.3 | **±0.2** |
| LLM | ±4.9 | **±1.0** |

## Controllability

(Compare Formula & Rules)

| | |
|---|---|
| gemini-2.5-pro: | 77% |
| ATA: | 73% |
| Miss. Clause | +4.7% |
| Miss. Predicate | +4.8% |
| Predicate Def. | +4.7% |
| ATA + **Corr:** | **87%** |
| with flash-lite: | 77% |

ATA ... uses gemini-2.5-flash without thinking
Base ... gemini-2.5-flash with built-in thinking

# Trustworthy[1] AI

- Accuracy & Stability
- Transparency & Explainability
- Human Agency & Oversight
- Privacy & Security
- Accountability
- Fairness & Non-Discrimination

# References

[1] Kowald, Dominik, et al. "*Establishing and evaluating trustworthy AI: overview and research challenges.*" Frontiers in Big Data 7 (2024): 1467222.

[2] Peer, David & Stabinger, Sebastian "*ATA: A Neuro-Symbolic Approach to Implement Autonomous and Trustworthy Agents*", arXiv:2510.16381, https://arxiv.org/abs/2510.16381

# Autonomous Trustworthy Agents (ATA)

Otera
David Peer
david.peer@otera.ai
**Machine Learning Researcher**