Verification Meets Uncertainty: Enabling Next-Generation Robotics, Logistics, and Compliance

Alvaro Velasquez,
Professor, University of Colorado Boulder,
Former PM, DARPA,
CEO, NSI
alvaro.velasquez@colorado.edu

September 22, 2026

University of Colorado Boulder

Neural vs. Neurosymbolic

THE STATE OF AI

Today's AI learns patterns *implicitly*—through massive data, models, and compute—making it inefficient and hard to trust.

OUR NEUROSYMBOLIC VISION

By *explicitly* embedding symbolic structures like physics and logic, AI can learn faster, generalize better, and be *mathematically verifiable* for trust.

University of Colorado Boulder

Verification: the next frontier revolutionizing AI4Math

nature

View all journals | Save

Explore content | About the journal | Publish with us | Subscribe

nature > news > article

NEWS | 18 September 2026

AI cracked the Navier–Stokes challenge. What does that mean for physics?

Physicists and mathematicians are going beyond the classic equations of fluid dynamics to understand turbulence – often with the help of AI.

By Nicola Jones

nature

View all journals

Explore content | About the journal | Publish with us | Subscribe

nature > news > article

NEWS | 22 May 2026

AI cracks 80-year-old mathematics challenge – researchers are astonished

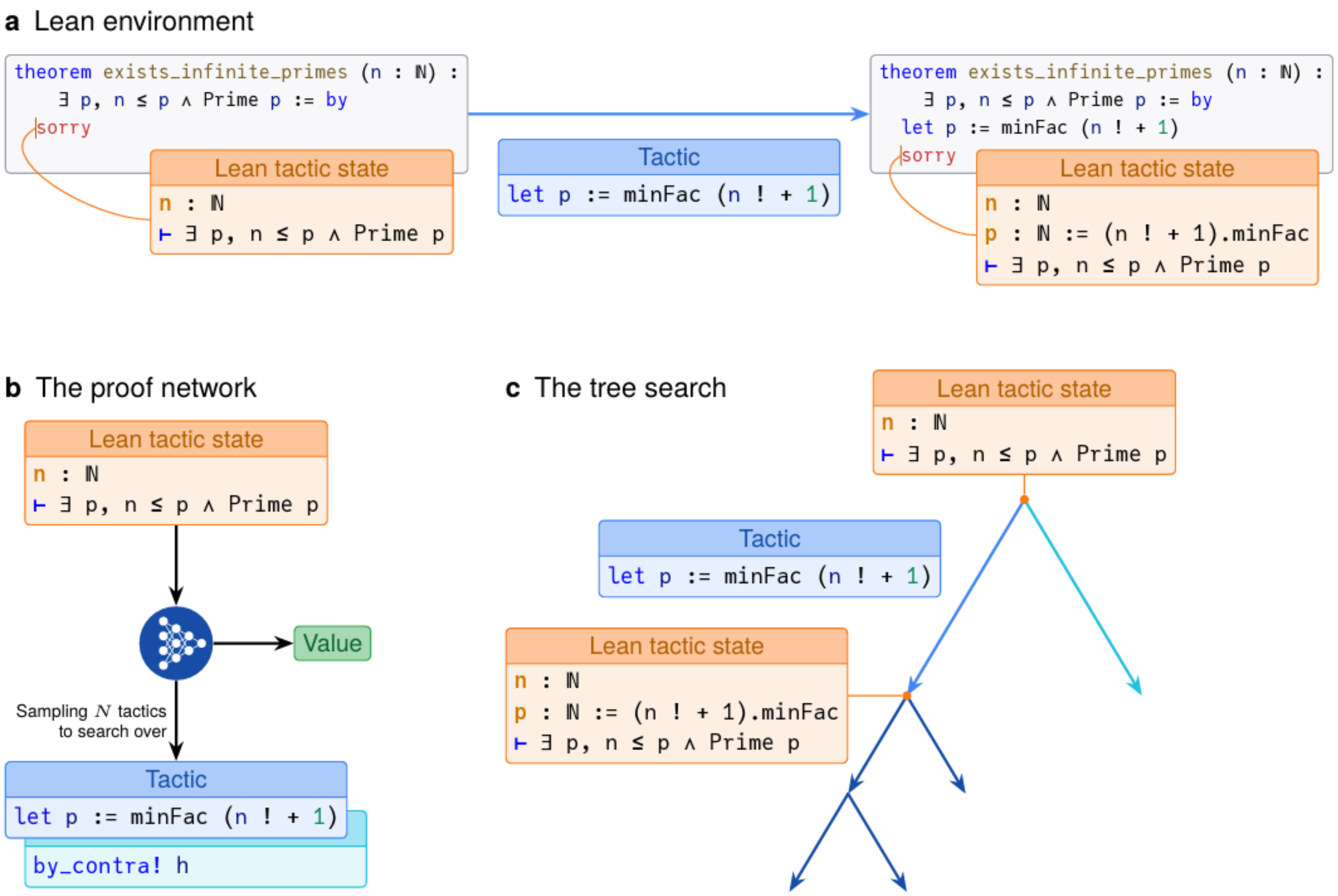
The late Hungarian mathematician Paul Erdős thought he had the last word on a geometry problem. Now an OpenAI chatbot has proved him wrong.

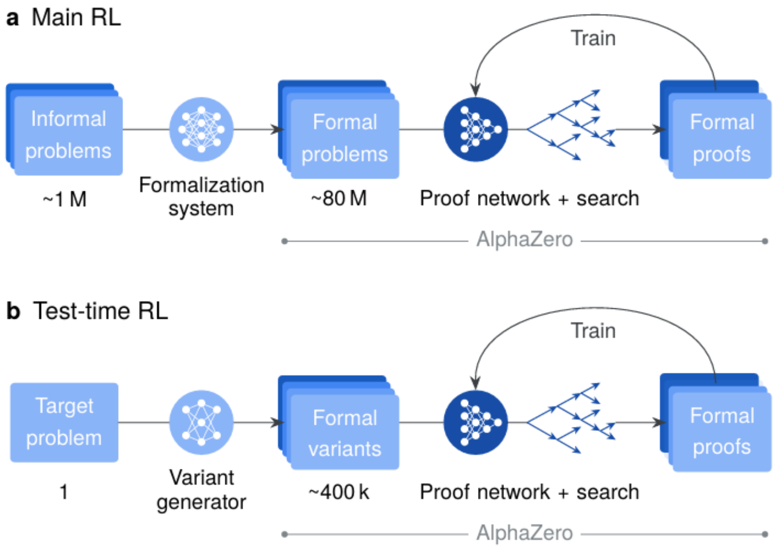
By Davide Castelvecchi

Verification: the next frontier revolutionizing AI4Math

University of Colorado Boulder

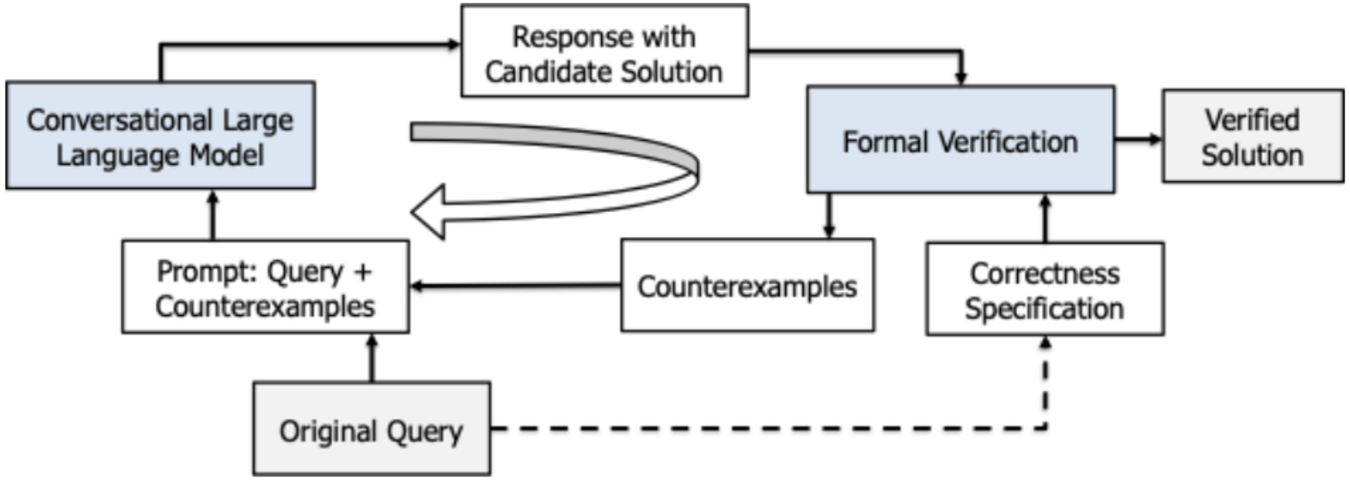
- Example: DeepMind's AlphaProof.





University of Colorado Boulder

Verification also works on AI4Code!



Jha, Susmit, Sumit Kumar Jha, Patrick Lincoln, Nathaniel D. Bastian, Alvaro Velasquez, and Sandeep Neema. "Dehallucinating large language models using formal methods guided iterative prompting." In *2023 IEEE International Conference on Assured Autonomy (ICAA)*, pp. 149-152. IEEE, 2023.

University of Colorado Boulder

But what about verification for systems subject to uncertainty?

Verification in the real world is subject to:

- Unknown models.
- Incomplete and small data.
- Limited operating picture.
- Ever changing requirements.

contested logistics

DDIL military systems

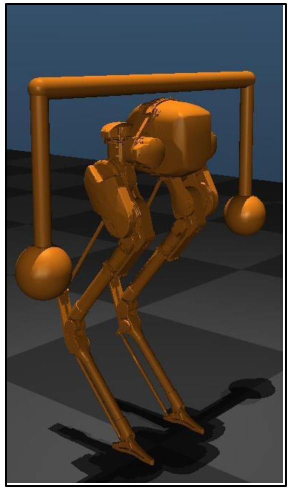
safety-critical robotics

molecular discovery

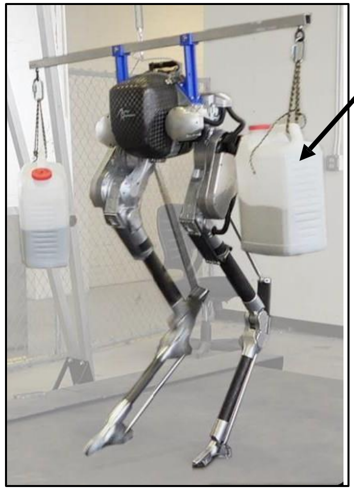
University of Colorado
Boulder

How bad can uncertainty be?

Containers



Simulation



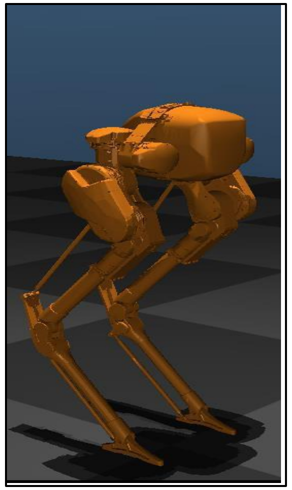
Real

Containers

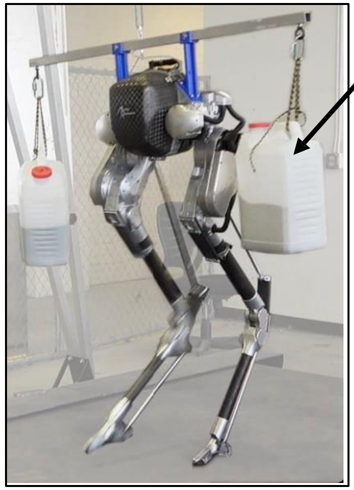
Low uncertainty: small sim-to-real gap

- Precise models for robot
- Precise models for containers
- Result: 96% success rate for achieving task

No
containers



Simulation



Real

Containers

High uncertainty: moderate sim-to-real gap

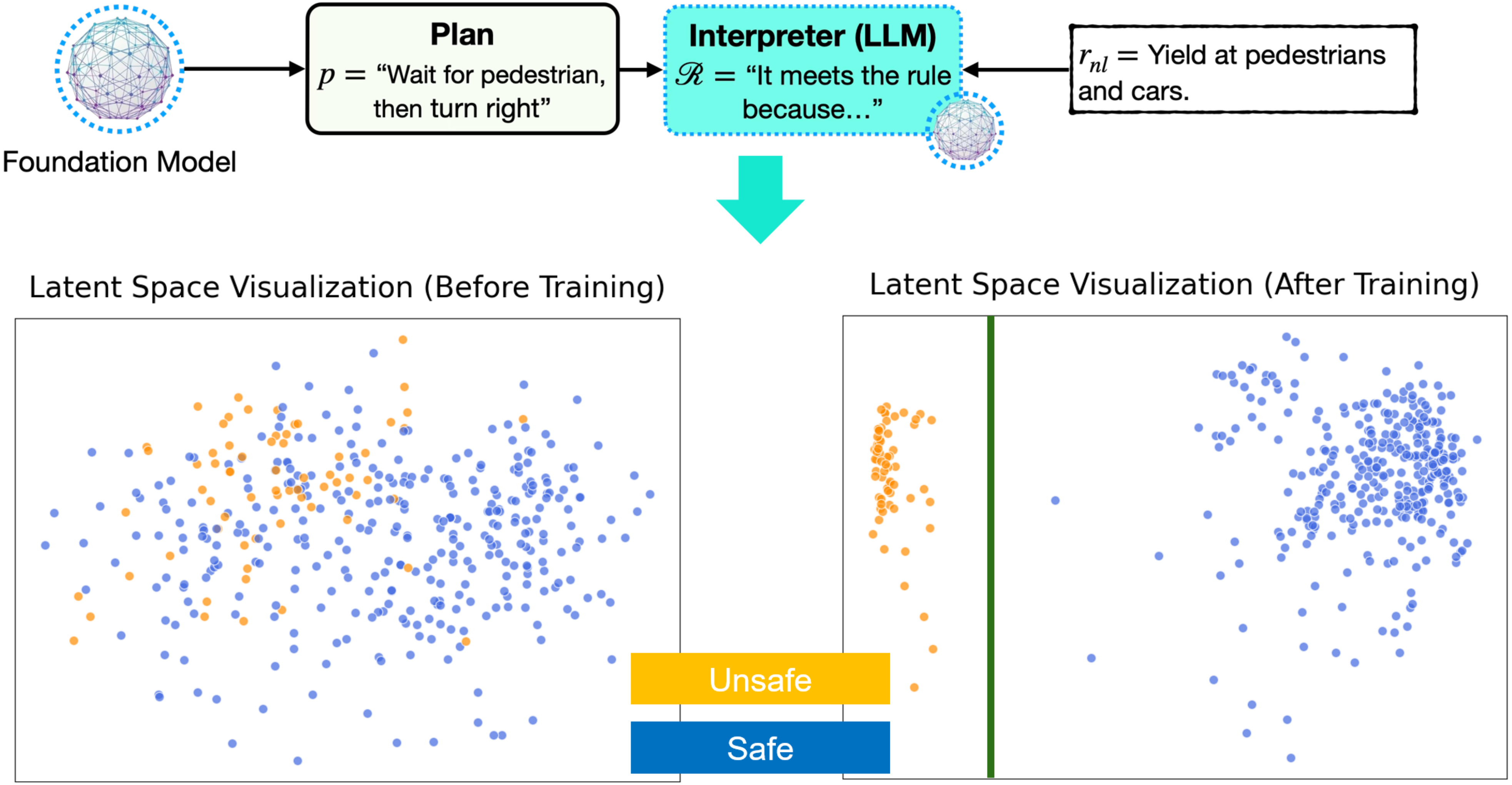
- Precise models for robot
- No models for containers
- Result: 0.1% success rate for achieving task

Conventional wisdom: scale up compute, parameters, data, etc.

- NVIDIA Project GR00T, Waymo autonomous driving, etc.

University of Colorado
Boulder

Verification under uncertainty is possible with a shared geometry



University of Colorado Boulder

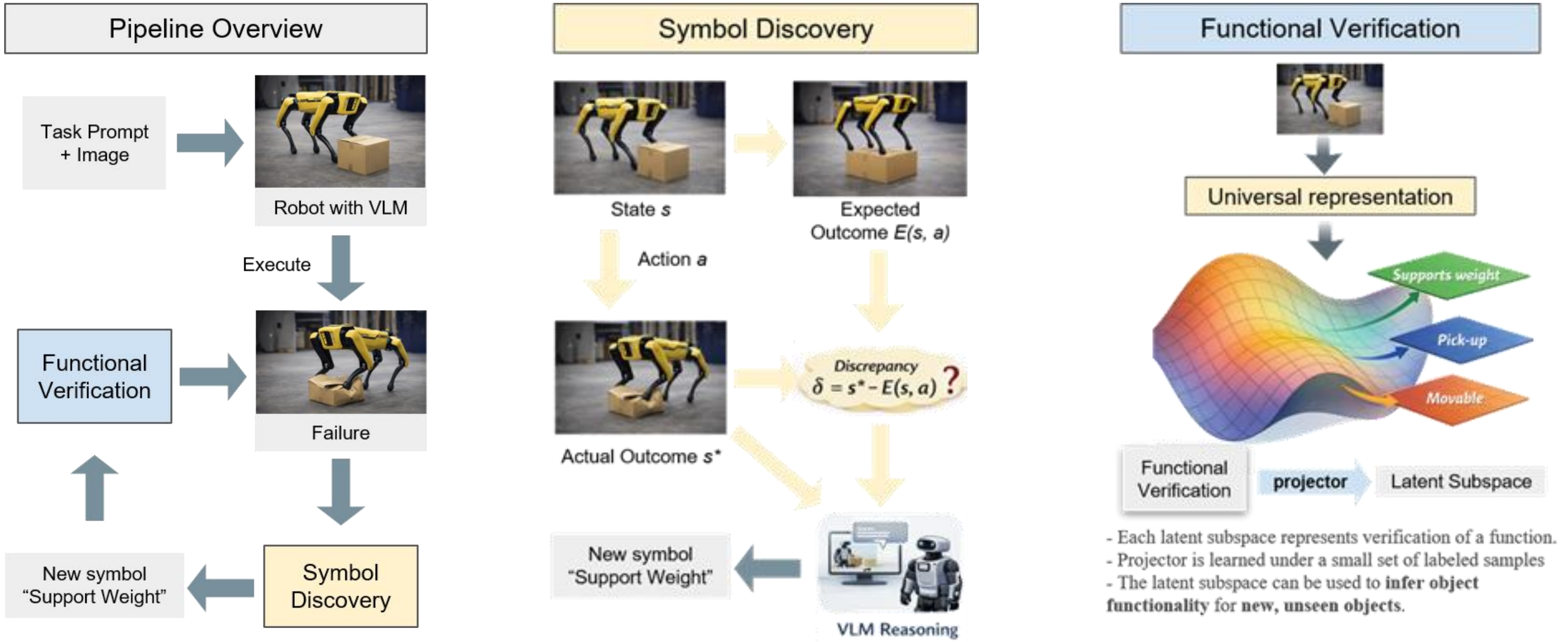
Verification under uncertainty is possible with a shared geometry

- Our methods enable verification under uncertainty for self-improving AI in settings where there is scarce and noisy data.
- We transform the verification problem into a prediction over the geometry of shared learned representations, enabling the successes of verification to scale to new verticals.
  - In robotics, uncertainty comes from our inability to model the real world and verification is transformed into prediction over learned robot sensor representations and the rules that should be satisfied.
  - In logistics, uncertainty comes from unforeseen supply chain issues and verification is transformed into prediction over learned representations of logistics plans and the routing constraints of the logistics problem.
  - In compliance, uncertainty comes from natural language interpretation of guidelines and verification is transformed into prediction over learned representations of a given system description (e.g., insurance claim) and the rules extracted from the guidelines (e.g., the Affordable Care Act).

University of Colorado Boulder

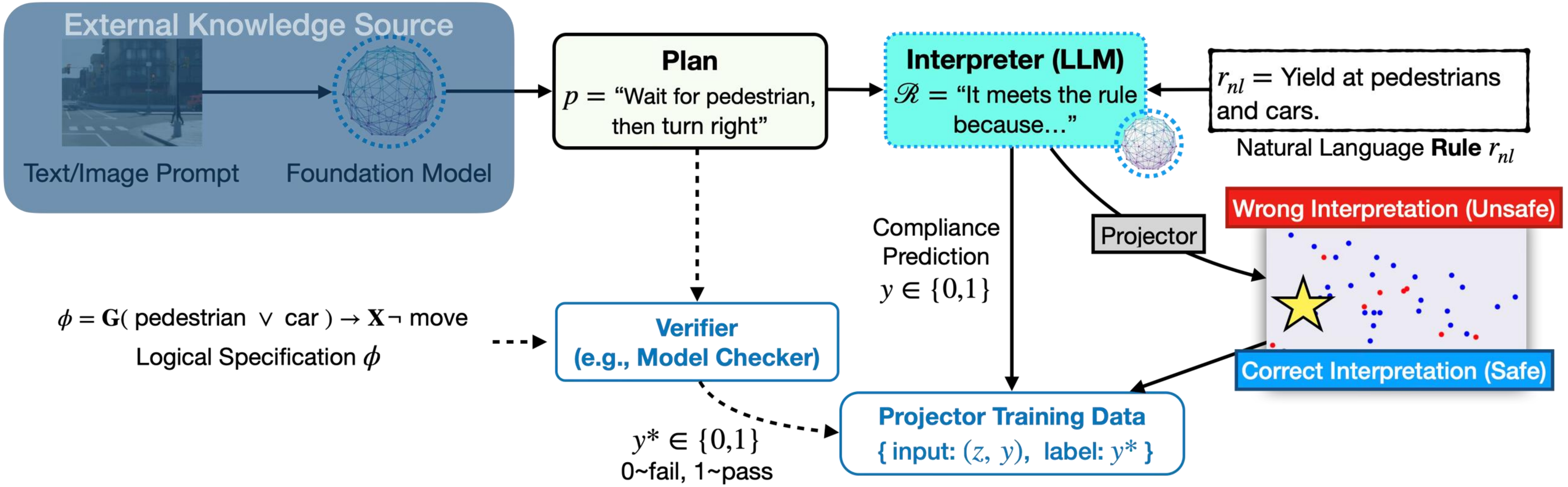
Use case: home robotics

Verification under uncertainty for the edge



Natural language verification under uncertainty

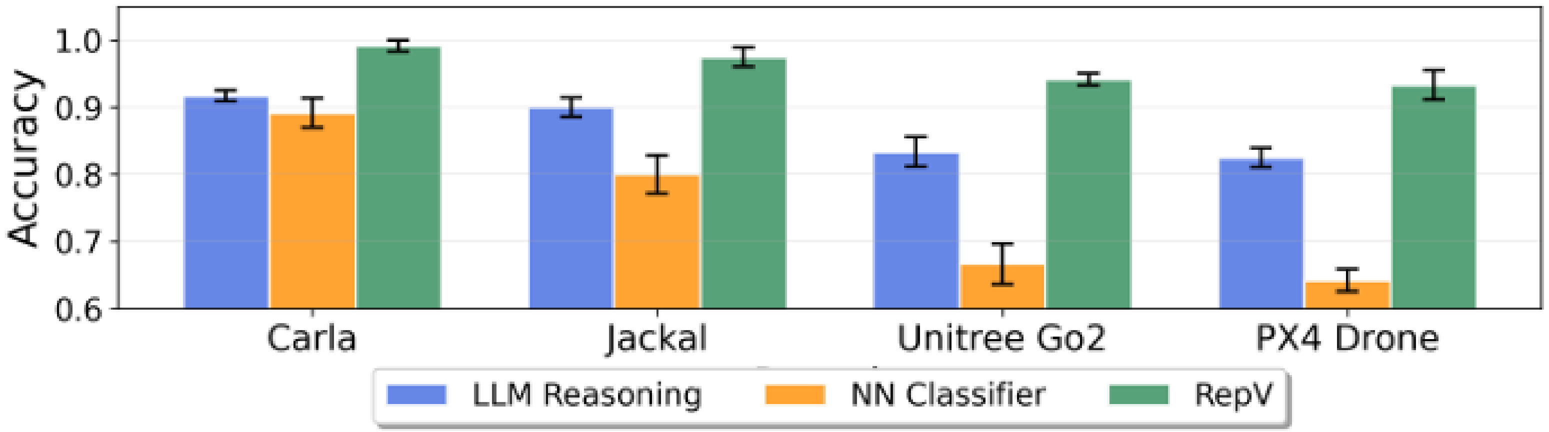
The projector constructs a safety-separable latent space, and predicting plan compliance against natural language rules is simply based on the spatial location of the projector's output representation for the natural language plan and constraints.



Yang, Yunhao, et al. "RepV: Safety-Separable Latent Spaces for Scalable Neurosymbolic Plan Verification." arXiv preprint arXiv:2510.26935 (2025).

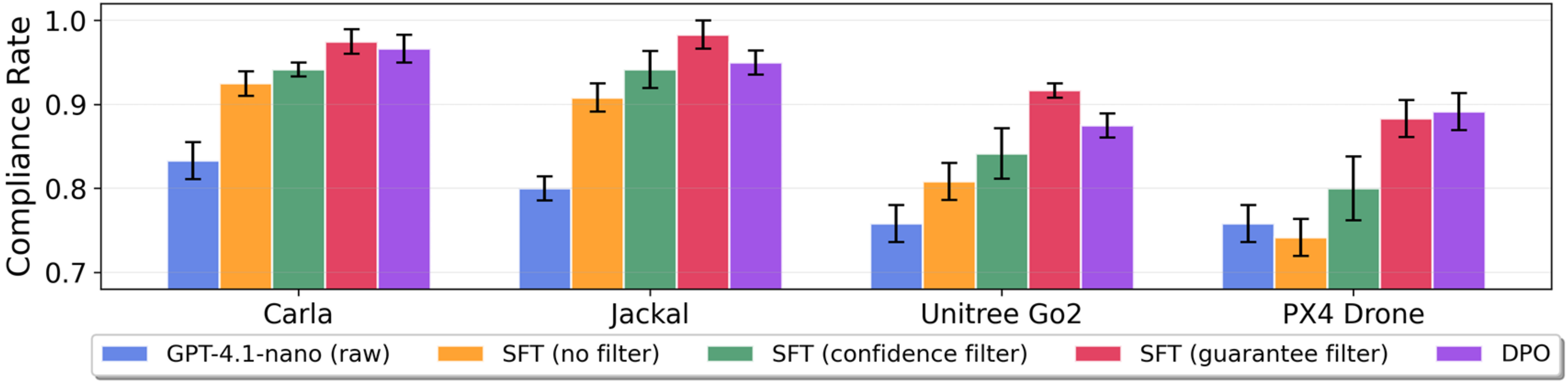
Natural language verification improves reasoning

Our approach outperforms powerful LLM reasoning baselines.



OpenAI GPT-4.1-nano

Using predictive verification for self-improving models

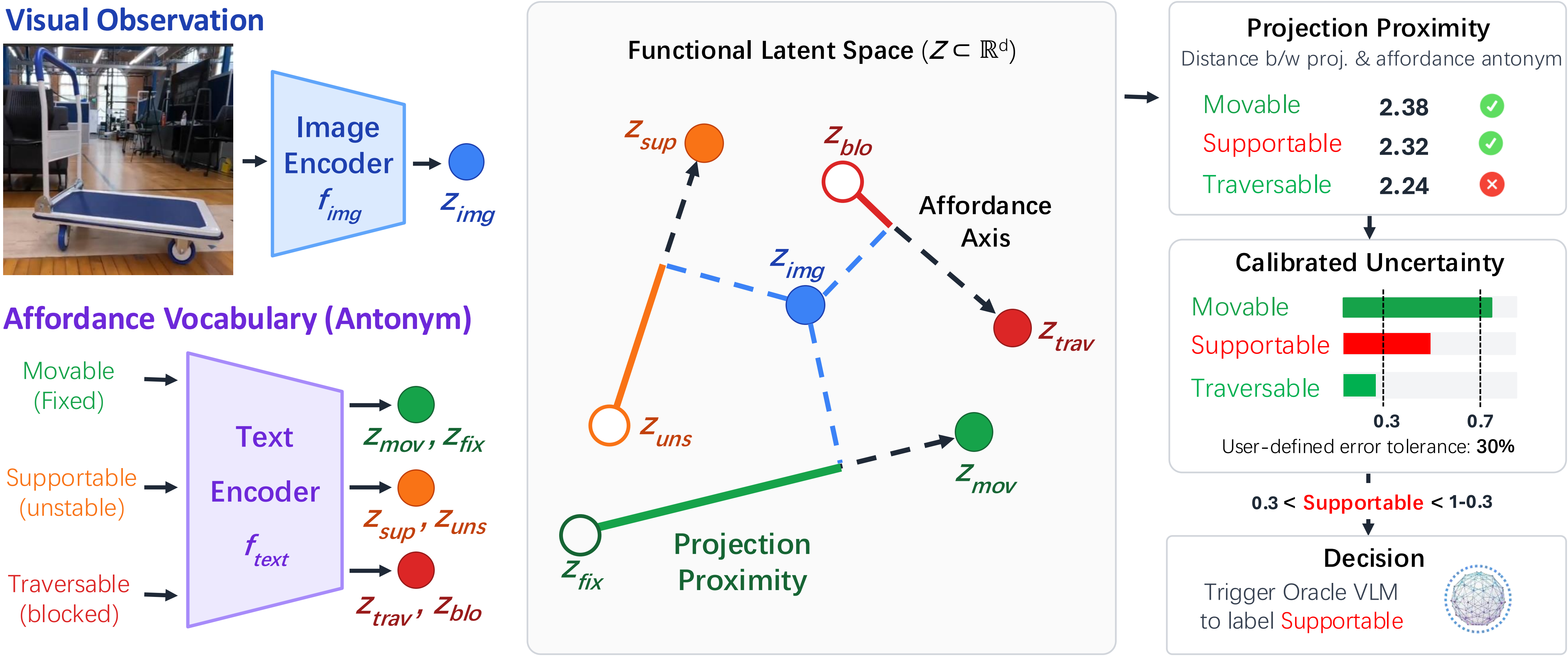


- SFT via Guarantee Filter (Ours): Fine-tuning restricted to samples satisfying the rules with probabilistic guarantee $p(y \mid z) \geq \tau = 0.8$.
- DPO (Ours): DPO using the probabilistic guarantee as a ranking signal between alternative plans.

University of Colorado Boulder

Using the shared geometry of uncertainty to discover function

Project Page



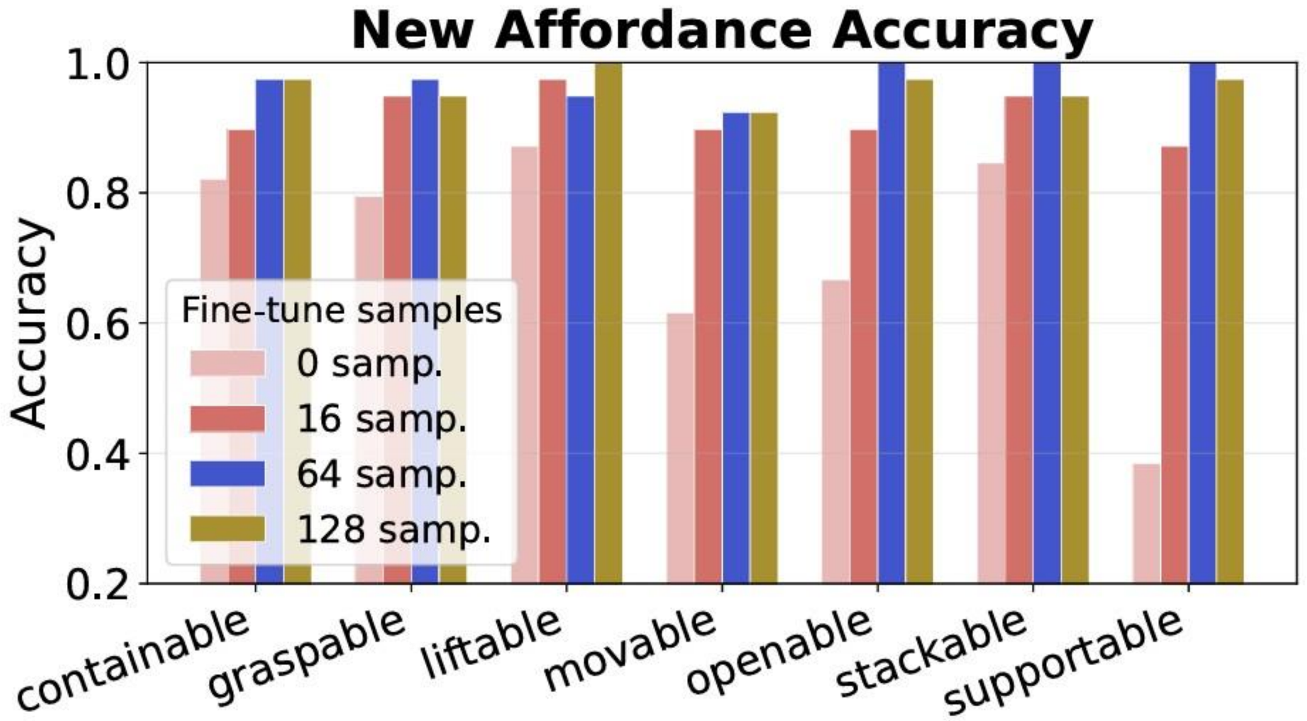
Siva, **Velasquez**, et al. "What Objects Enable, Not What They Are: Functional Latent Spaces for Affordance Reasoning." *Conference on Robot Learning* (2026).

Using the shared geometry of uncertainty to discover function

Given input image and task prompt, **infer** object affordances relevant to the task and estimate associated inference **uncertainty** in **real time** while **discovering** new affordances when required.

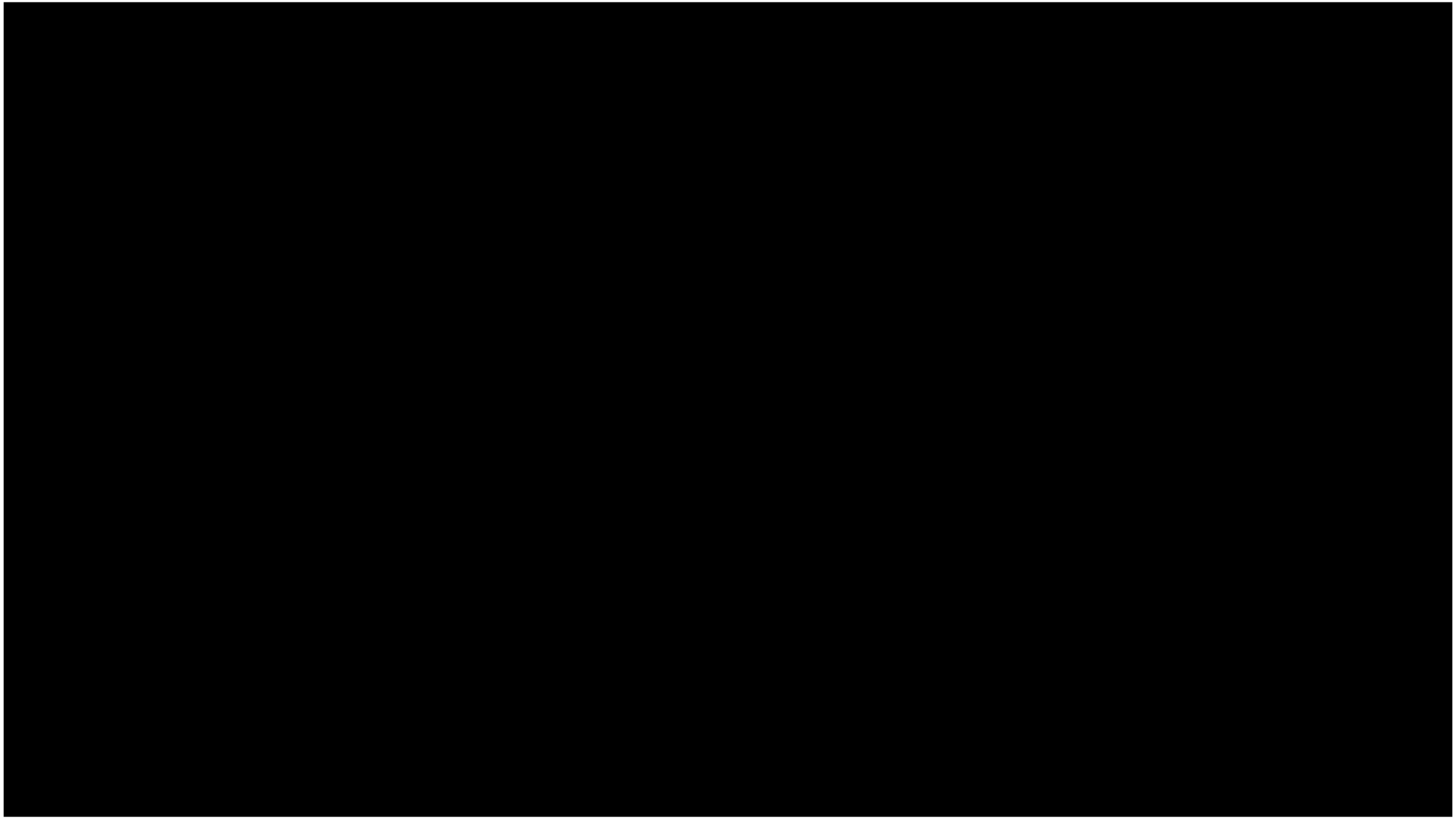
Proof of concept: Cloudless low-power edge robotics.

- **Small data:** <200 data points required to learn and reason about the function of objects with greater accuracy than GPT-5.4 (94% versus 79%).
- **Real-time inference:** 100x faster inference when compared to GPT-5.4 (22ms vs. 2.6s)
- **Lightweight:** 150M parameters compared to 2.2T parameters in GPT-5.4, 20B in GPT-OSS.
- **Improved loitering time:** 15% longer duration compared to a GPT-OSS-20B model running on a SPOT robot.



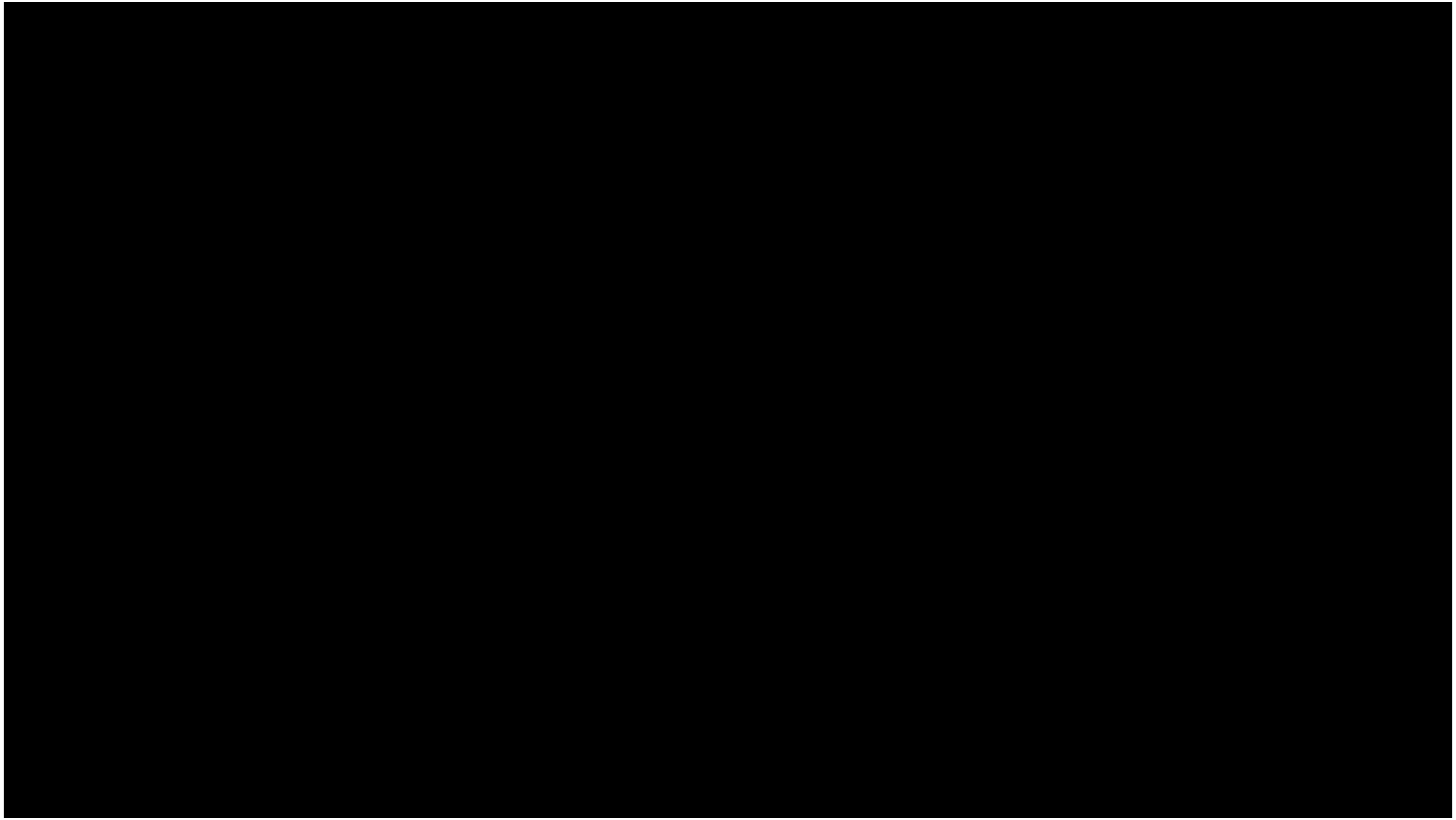
Siva, **Velasquez**, et al. "What Objects Enable, Not What They Are: Functional Latent Spaces for Affordance Reasoning." *Conference on Robot Learning* (2026).

Use case: crater exploration for lunar prospecting



University of Colorado Boulder

Use case: crater exploration for lunar prospecting

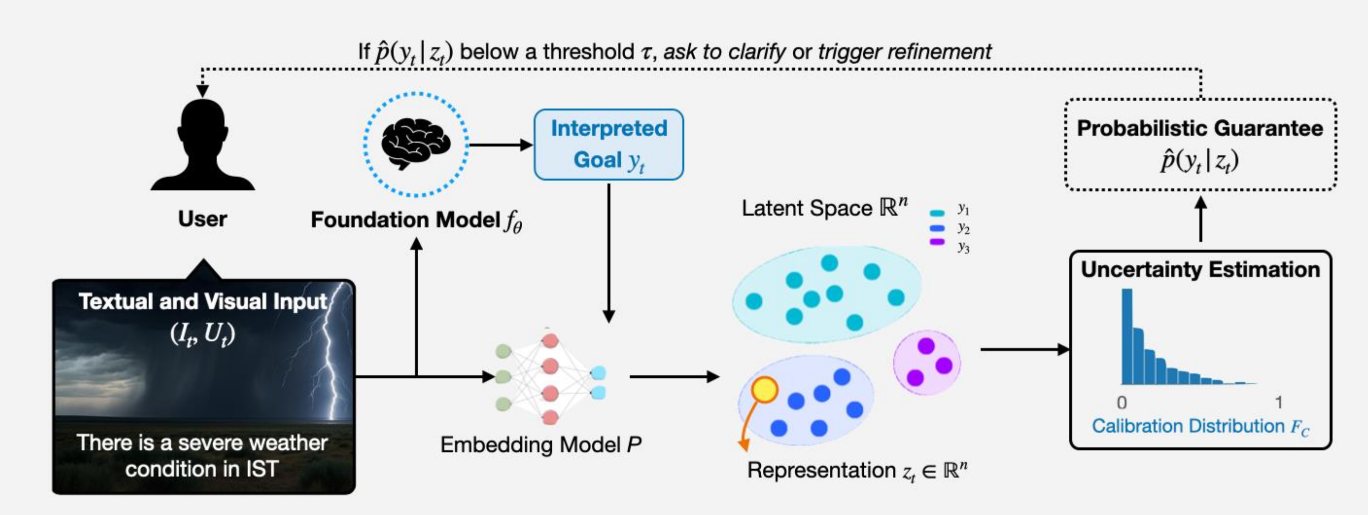


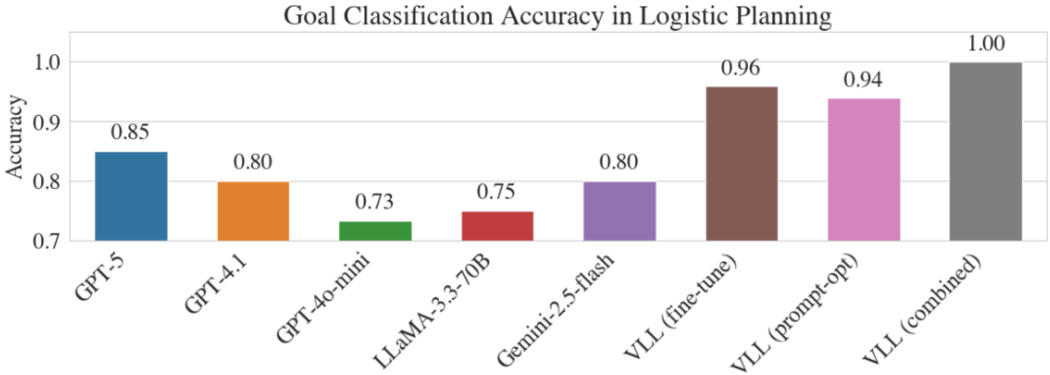
University of Colorado Boulder

Other applications: logistics and compliance under uncertainty

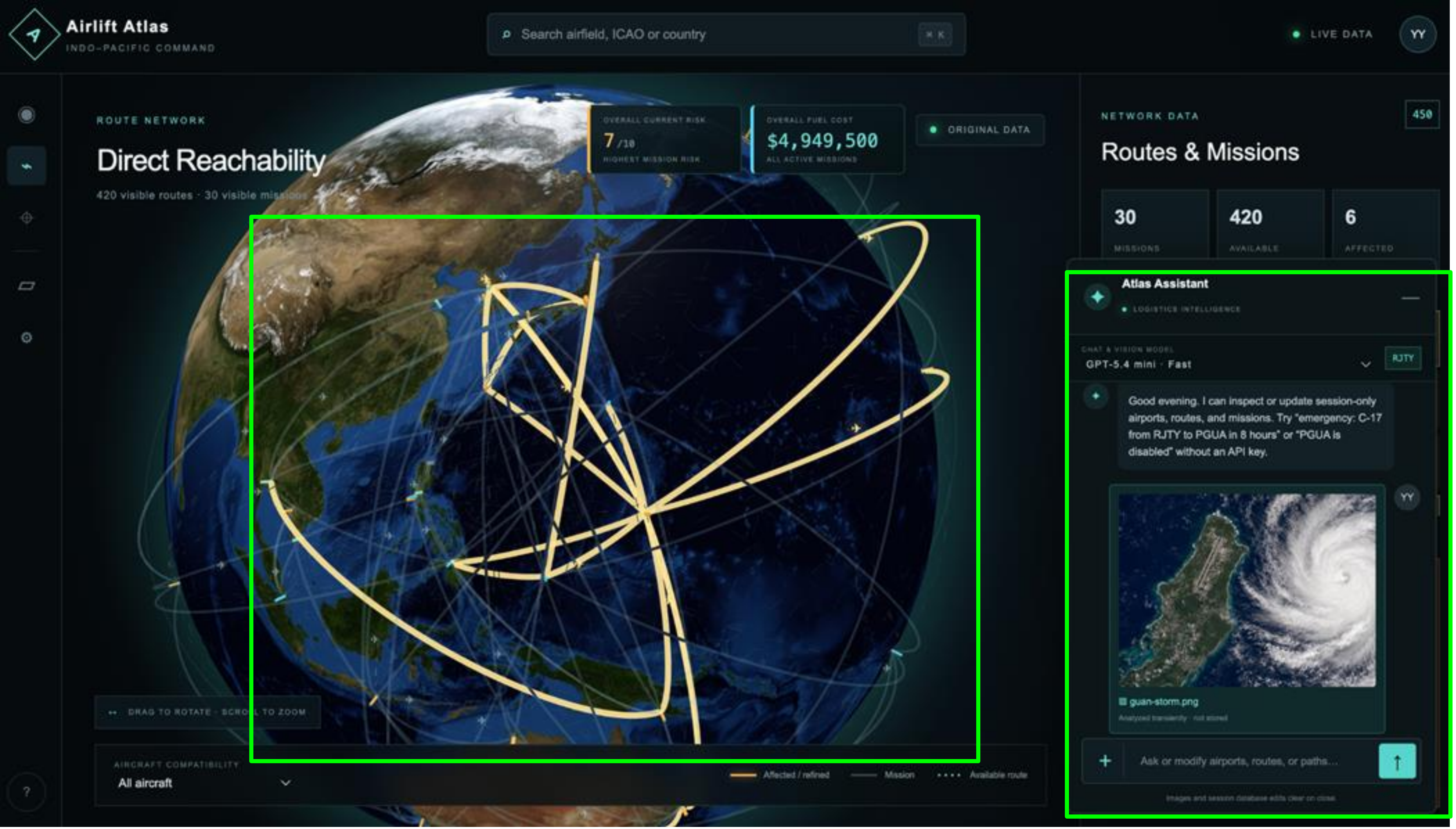
Logistics under uncertainty

- We have developed an architecture that outperforms GPT-5 for military logistics that satisfies constraints from natural language and visual data.
- Inputs are converted into structured goals and dispatched to symbolic planners. Predictive verification under uncertainty is performed on the plans produced.



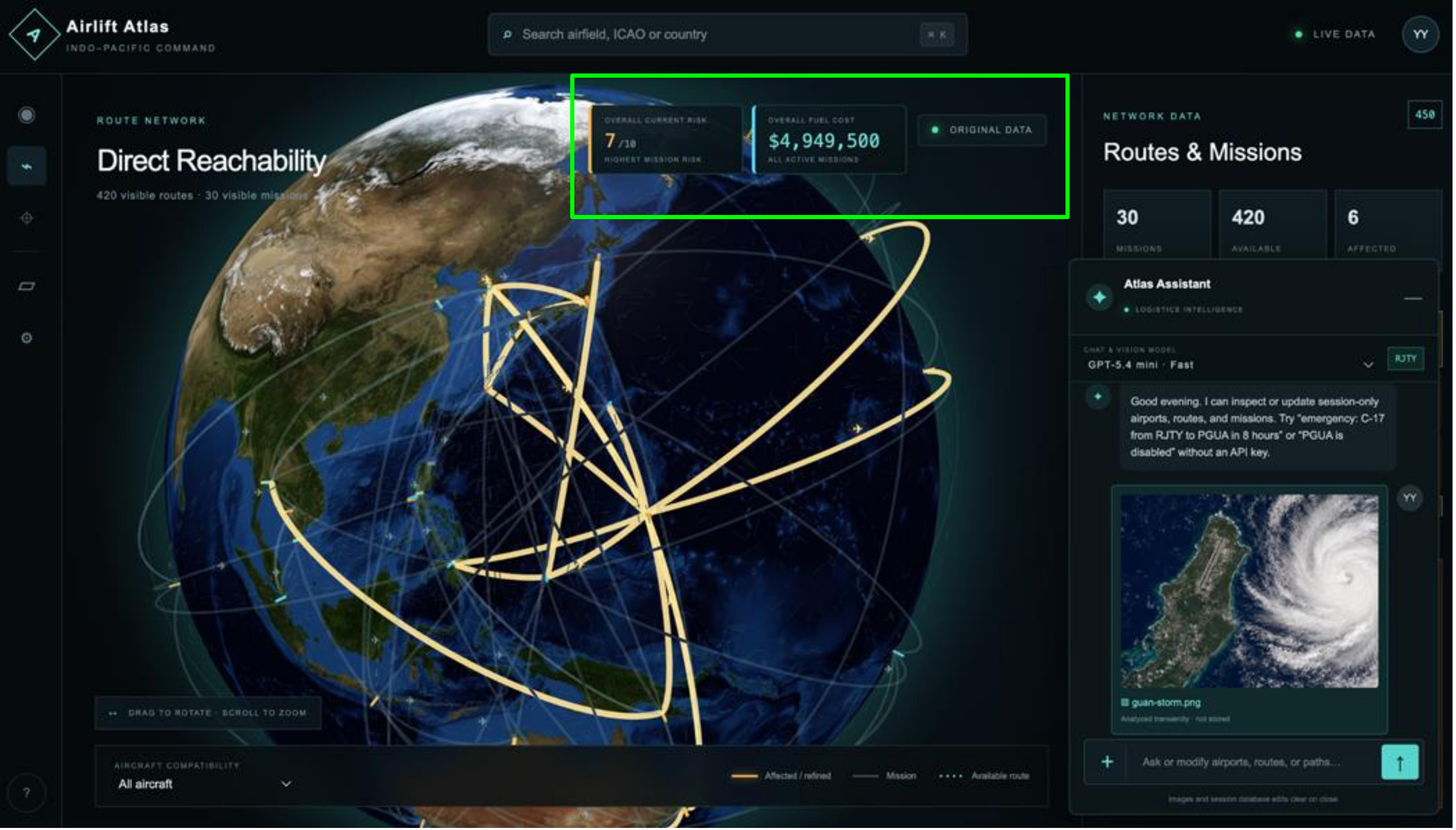


Yang, Yunhao, Neel P. Bhatt, Christian Ellis, Alvaro Velasquez, Zhangyang Wang, and Ufuk Topcu. "Foundation Models for Logistics: Toward Certifiable, Conversational Planning Interfaces." NeuS 2026.

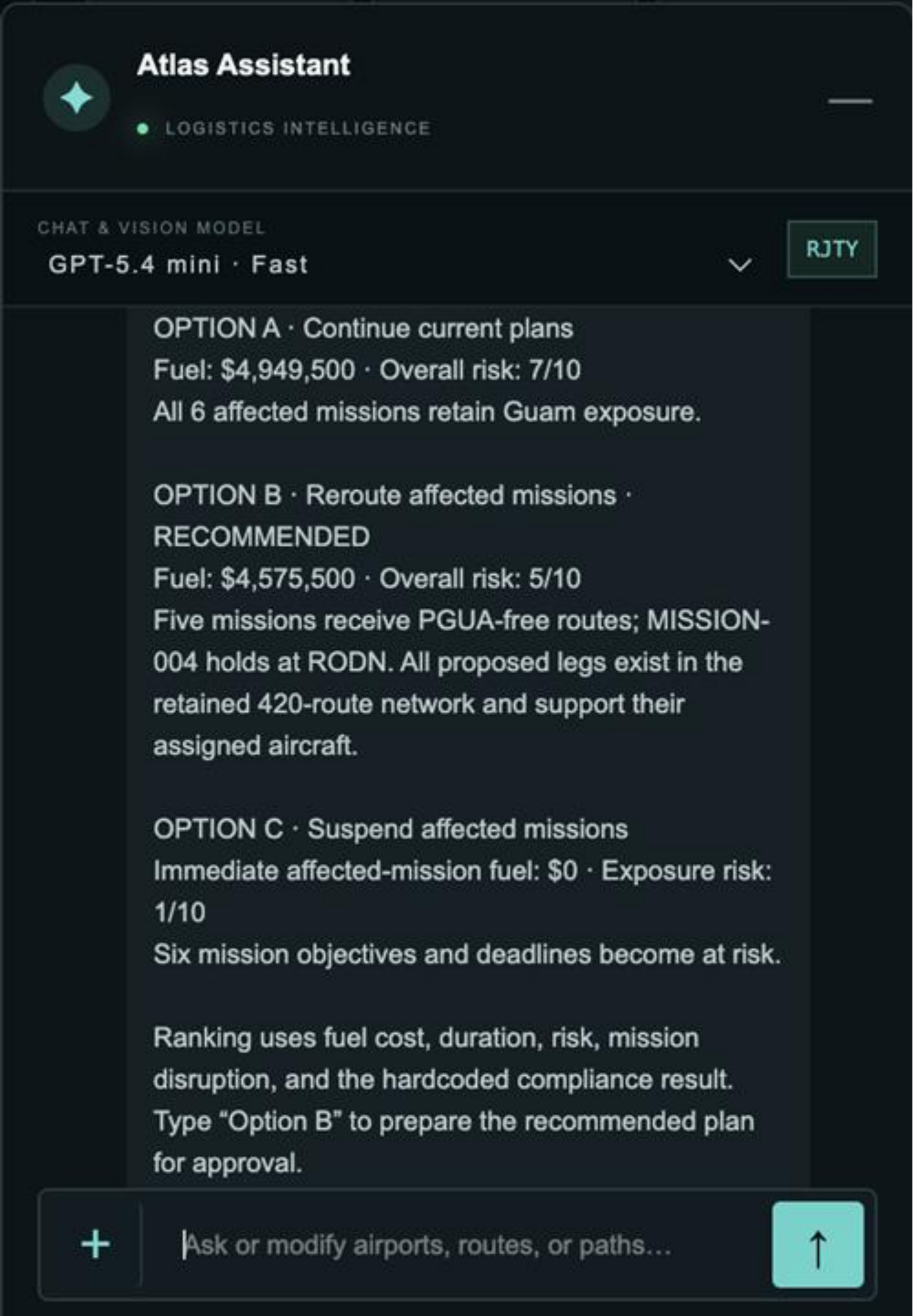


user upload image to the chat

The VLL determine which agents (paths) will be affected

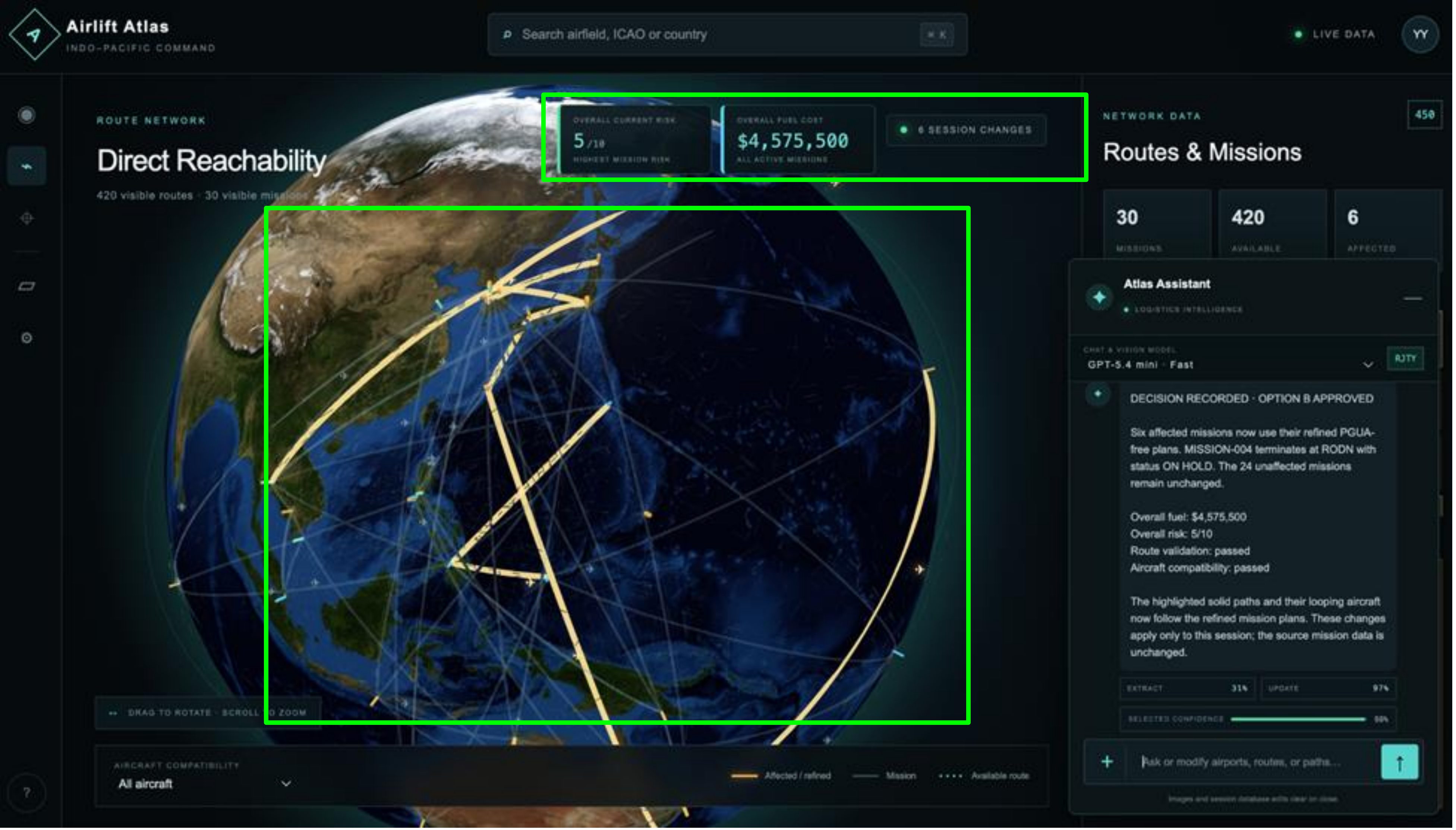


Keep track of the fuel costs and risks (calibrated probability of compliance)



The VLL applies multi-agent path planning to the impacted paths and provide three options with **explanations**

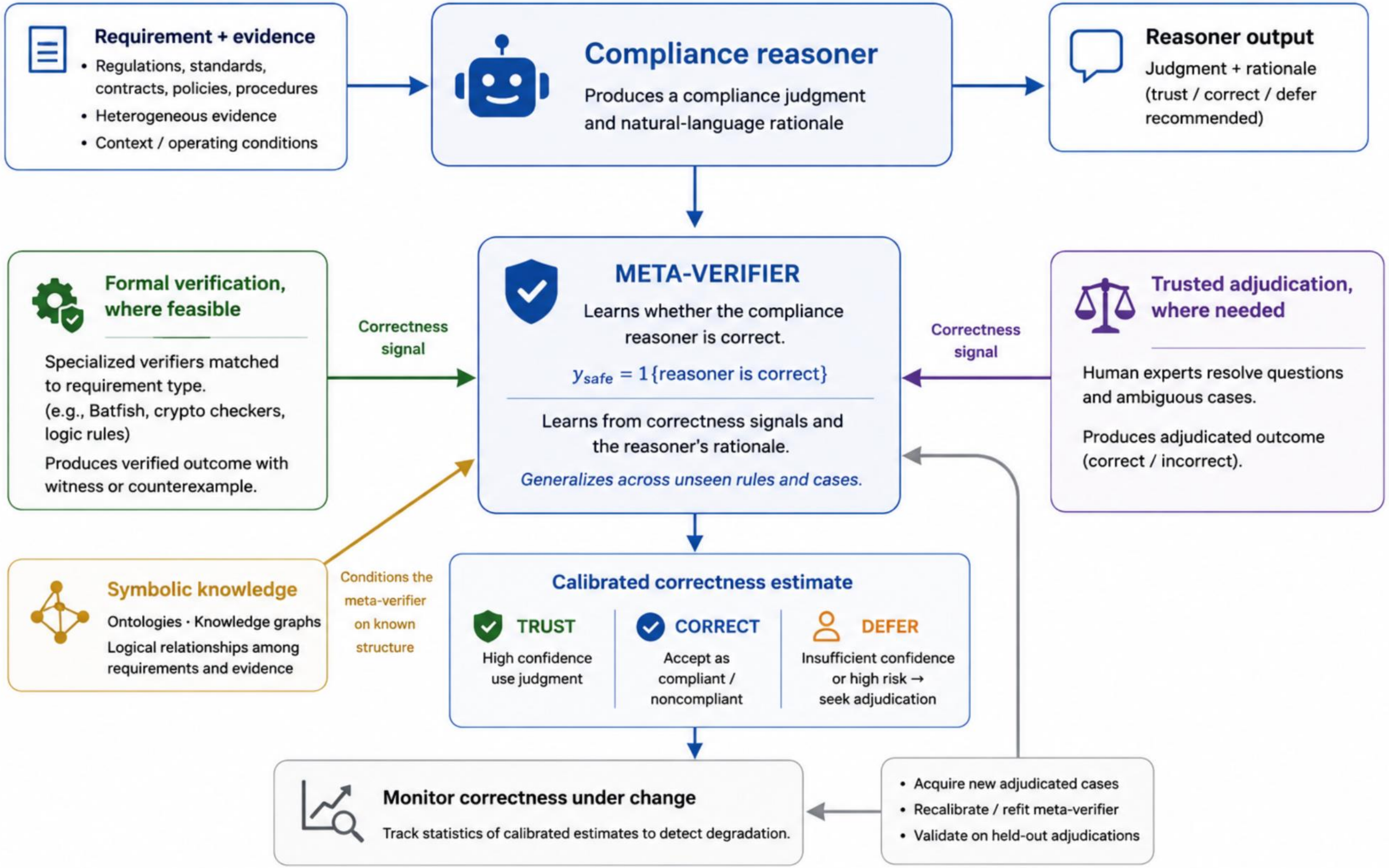
It then **ranks the options** based on the costs, risks, mission completion rate, etc and gives user a recommendation



Risk and cost are updated accordingly

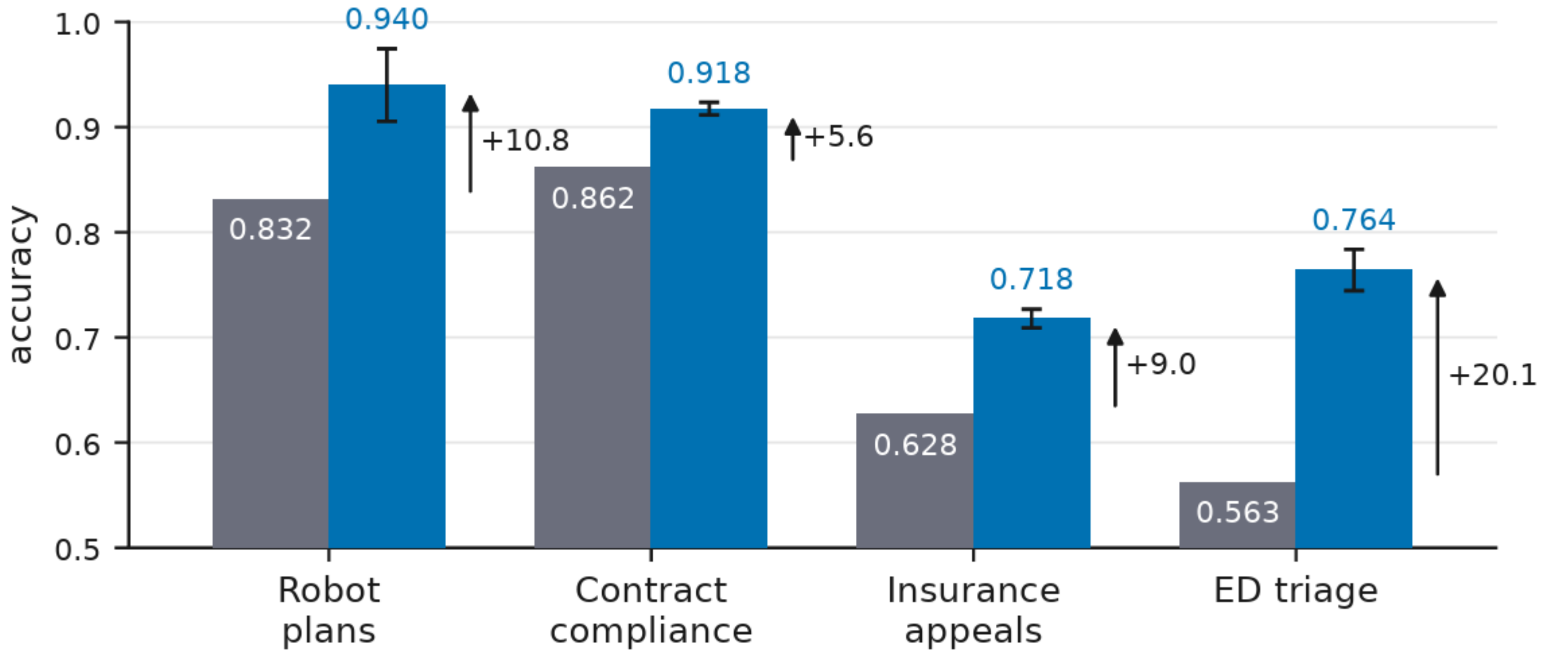
The VLL update paths once user makes the decision

Compliance under uncertainty



Compliance under uncertainty: medical use case

Insurers deny roughly 17% of in-network claims — about 48 million a year in the ACA marketplace alone — and fewer than 2000 denials are ever appealed. Yet when someone does appeal, about 41% of denials get overturned. Nobody re-examines the other 99.8% — not patients, and not the hospitals sitting on the money. That informs the market opportunity: California publishes 42,000+ adjudicated appeal decisions, so we can rank a hospital's denial pile by calibrated probability of overturn and have the same appeals staff file the winnable, high-dollar ones first.



https://www.kff.org/private-insurance/claims-denials-and-appeals-in-aca-marketplace-plans/

University of Colorado Boulder

Conclusion

- Verification is revolutionizing AI4Math
- For verification to do the same for robotics, logistics, compliance, and other domains, it must handle uncertainty
- We achieve verification under uncertainty by mapping systems and uncertainty to a shared geometry where we can compute confidence of prediction and the amount of additional data required to achieve a threshold of confidence

Thanks!

alvaro.velasquez@colorado.edu