Atlas Computing

AI advances in different domains will lead to:

Domain	Benefit	Risk
Systemic	Improve human systems	Outpace human oversight
 Cyber	→ Fix software vulnerabilities	! Make new computer viruses
 Bio	→ Cure diseases	! Make new pathogens
 Law	→ Patch legal ambiguities	! Find legal loopholes
 Society	→ Reduce information overflow	! Create disinformation

Everyone wants human-level AI agents.

No one can define “good behavior for humans.”

“Good AI” is...

→ harder,

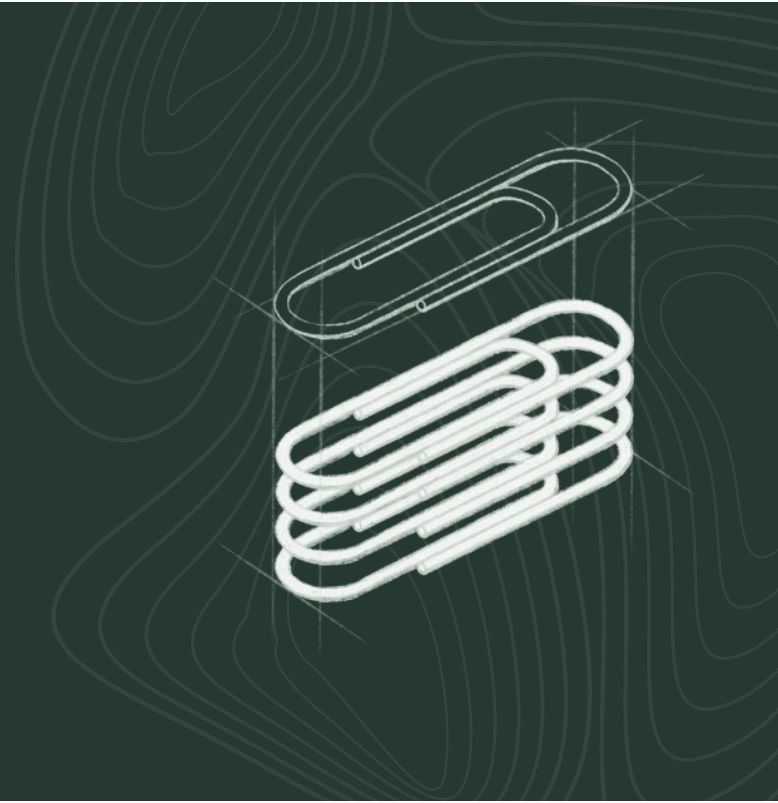
→ subjective,

→ a moving target

Today's oversight doesn't scale to everyday AI agents.

When it cost <$1/day/"person" to simulate an "employee", how do you understand or steer their activities?

The resulting arms race of "wrangling the most AI agents" could destabilize any (and thus, every) human system.

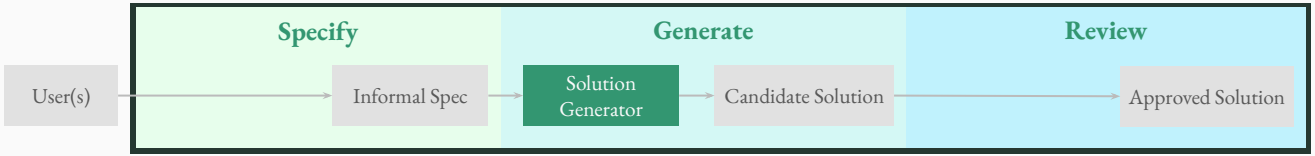


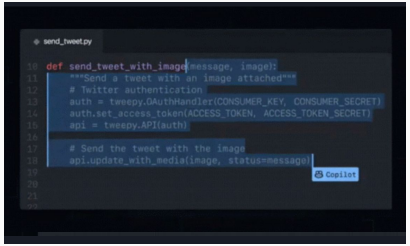
Make tools to set rules
that the AI proves it's obeying.

AI agents should not be monolithic “black boxes”

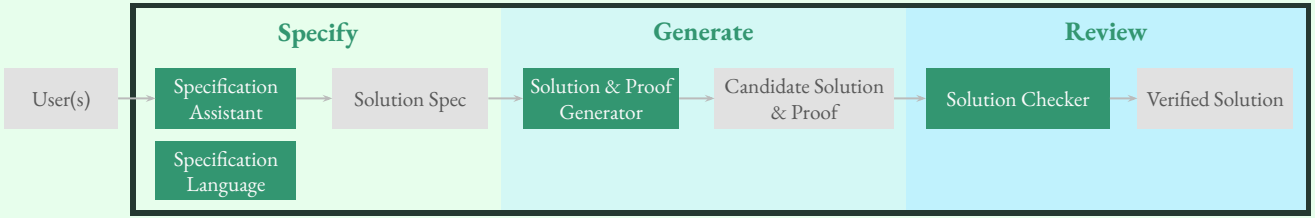
AI should create intermediate outputs for humans to check.

Not this





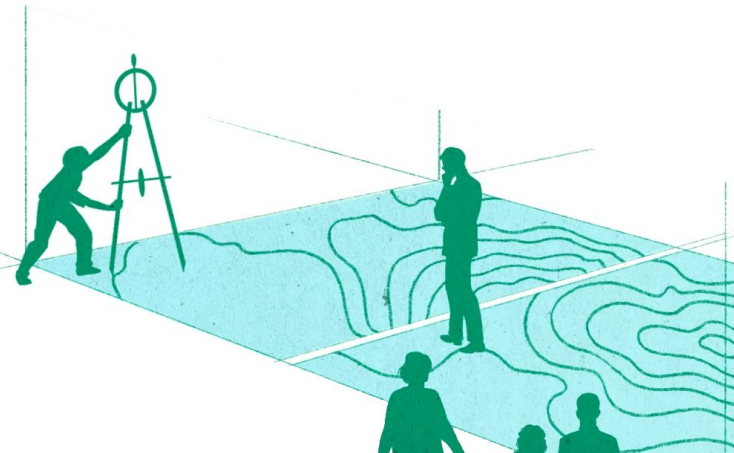
But this



Enable scalable human review

An international standards organization for verifiably governable AI

Make it easy and rewarding
for individuals, companies,
and governments
to build specification-driven AI



Help ✅ each of these in high-leverage ways

(e.g. match great people to funding, prototypes, customers, and investments)

✅ Done
🚧 We're building
➡️ SOON Others' progress

Stage	Cyber	Bio	Law	Society	etc.
Relevant experts aware of AI risk?	✅	➡️ SOON	➡️ SOON	➡️ SOON	
Specification language exists?	✅	🚧			
AI enhanced spec generators exist?	🚧	➡️ SOON			
AI enhanced solution generators exist?	➡️ SOON	➡️ SOON			
AI enhanced solution checkers exists?	➡️ SOON				

Progress so far
Atlas was founded Q4 2023

Domain 1

Cyber

→ **Goal**
Solve "last-mile problem" for proofs with existing spec languages

→ **First Deliverables**
Formal Methods + AI 2-pager and Coq to Lean translation

→ **Potential Second**
Prototype natural language to formal spec conversion

Domain 2

Bio

→ **Goal**
Make a new spec language

→ **First Step**
Designed and refining proposal for a toxicity forecasting competition

Building a

Community

→ Co-organized 2 workshops, organizing 3 more events
See our events page

→ Organized an email list
Organizing discussions among collaborators

AI risk is better averted one domain at a time.

Rather than attempting to solve the whole problem at once.

LLMs can scale Formal Verification.

(an existing but costly specification language for software)

The Team

CEO

Evan Miyazono

Protocol Labs* → Head of Network Goods
Created and led a venture studio (up to 25 people; ~$7M/yr), and spun out 3 for-profits & 3 non-profits:

Open Source Observer · Discourse Graphs · hypercerts · Funding the Commons · Gov4Git · OPEN IMPACT FOUNDATION

→ Head of Research
Created and led the research grants program, metascience, and special projects teams.

Caltech PhD → Applied Physics

Stanford BS/MS → Materials Science

Software Lead

Daniel Windham

Apogee Research → Principal Software Engineer
Co-led software development and usability on STITCHES, one of the most successful DARPA program results of the decade.

Coda → Software Engineer
Shipped 13 projects in 22 months to support pre- and post-launch growth; Coda now has 1M+ users.

Harvard BS → Physics, Computer Science

* First PhD hire at now one of the most respected companies in web3

The Team: Why?

Prior to joining ARIA (UK's new ARPA-like entity)

David "davidad" Dalrymple proposed the AI architecture that Atlas is pursuing while a researcher on Evan's team.

Prior to co-founding Atlas Computing

Evan spent 6 years funding research and building teams to launch products based on davidad's ideas, like Filecoin's Proof of Replication and Hypercerts

Prior to co-founding Atlas Computing

Daniel spent the last 5 years transitioning DARPA technology for integration and adaptation of heterogeneous Air Force systems

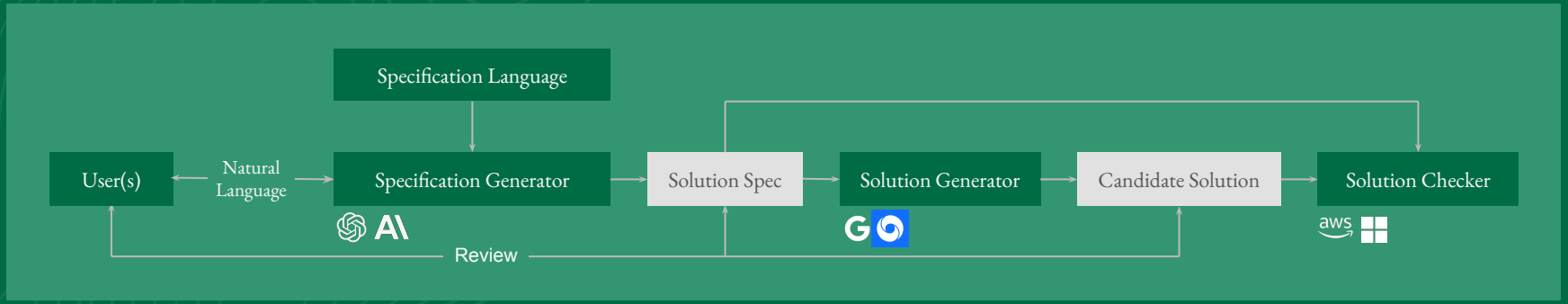
- → AI Experts
  - ARIA Programme Director
  - CEOs of AI and tech companies
  - MIRI board member
- → Governance, Deliberation & Ethics researchers
  - Founders and researchers
- → Venture Capitalists & other CEOs
  - MacArthur Fellow & labor organizer
  - Founder / Investors

This architecture is the
best way to govern AI,
but humans only stay in control
if everyone uses it.

As a non-profit, we can welcome AI companies into an interoperable ecosystem instead of competing for market share.

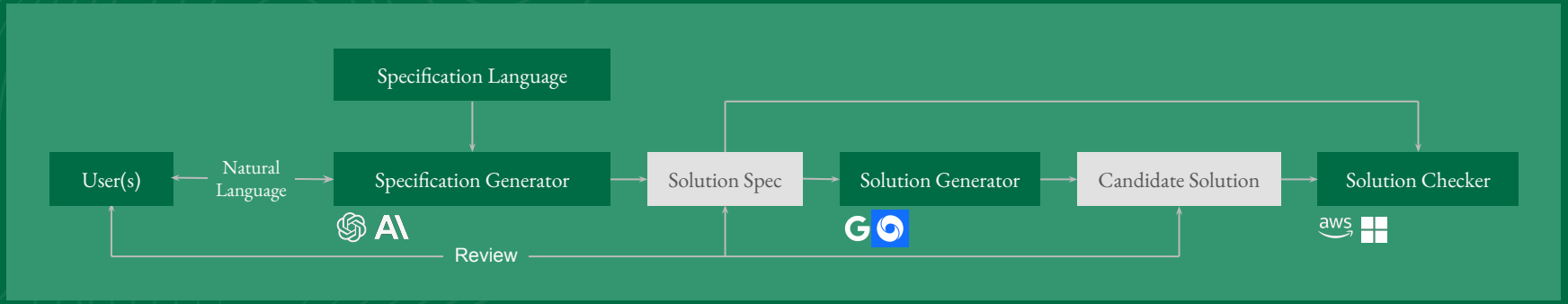
- → We recruit and support companies & researchers
- → We make sure the IP is available
- → We do the work no one else will

People will like it if they see themselves in it...



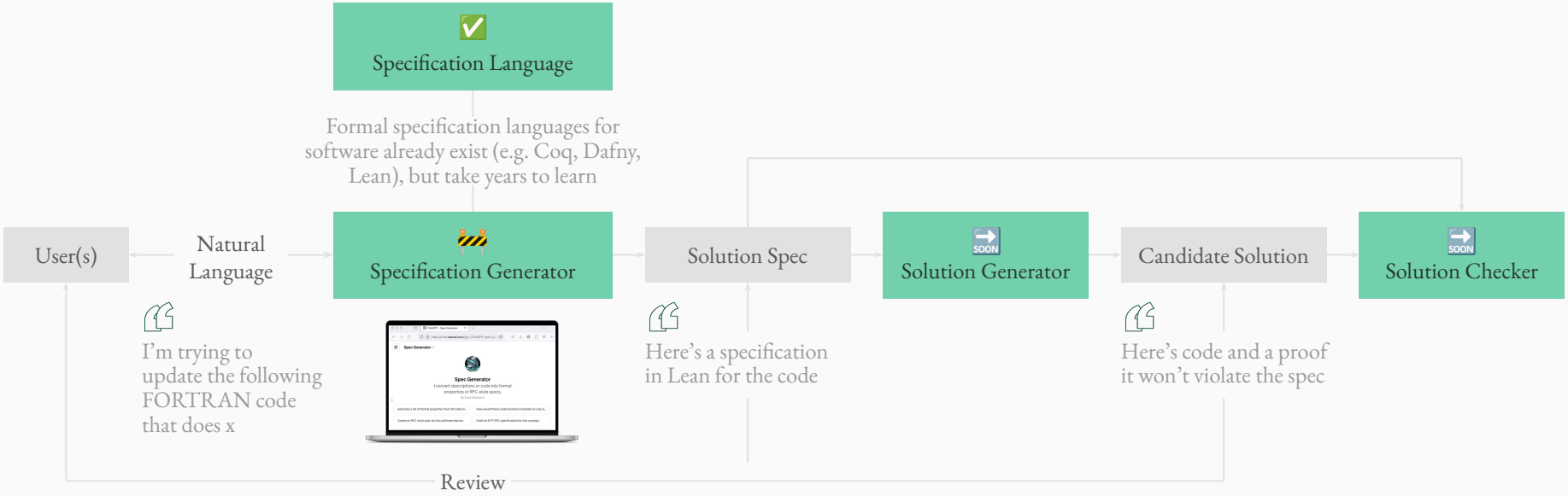
- → We recruit and support companies & researchers
- → We make sure the IP is available
- → We do the work no one else will

People will like it if they see themselves in it...



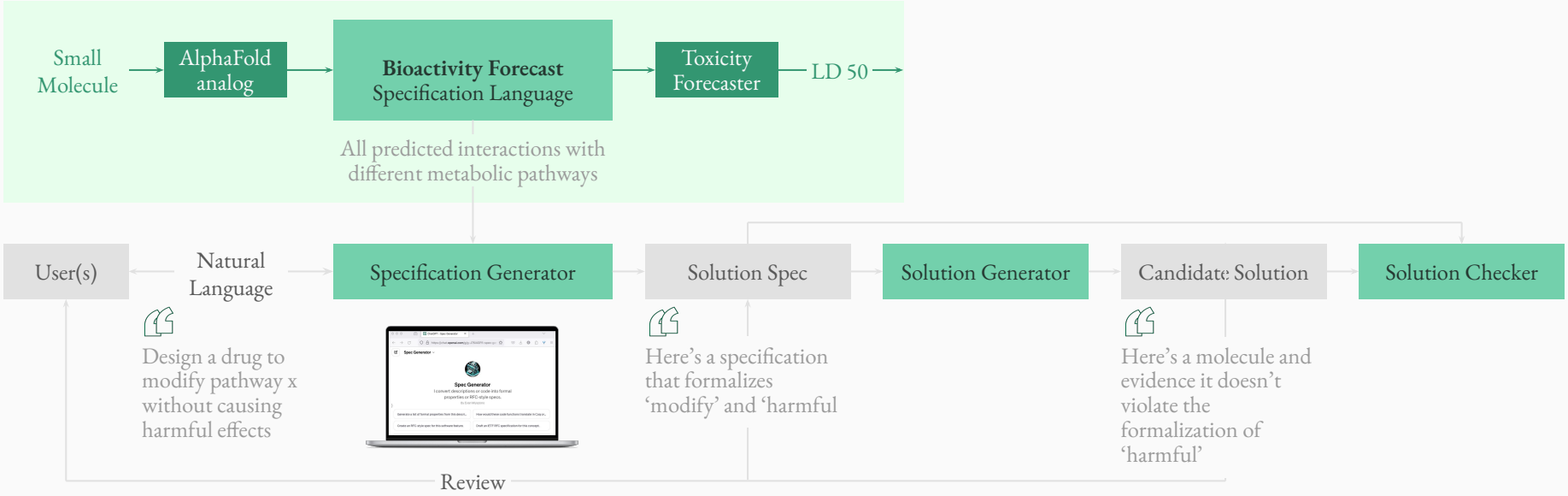
Updating software is critical but risks breaking capabilities

Specs make this update secure (and future updates easy) - read more here



Organize a competition to create tools to predict bioactivity & toxicity

You can now screen new molecules with predicted bioactivity. This could automate drug discovery, environmental impact, or similarity analysis for controlled substances



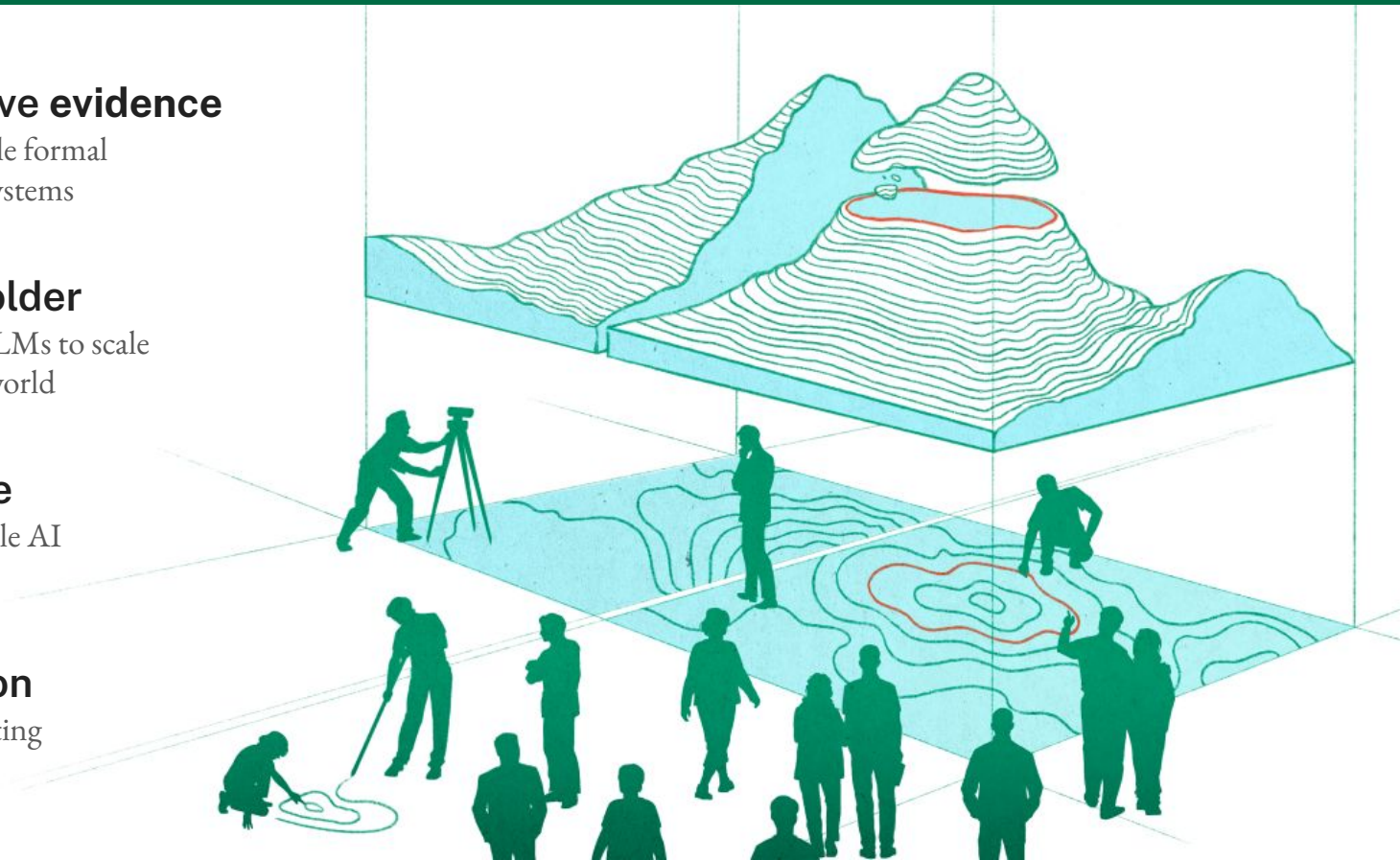
Selected EoY 2024 goals

→ **Produce persuasive evidence**
that LLMs are ready to scale formal verification in real-world systems

→ **Identify a stakeholder**
who by Q3 who can use LLMs to scale formal verification in the world

→ **Host a conference**
for 100+ people on provable AI safety properties

→ **Start a competition**
to advance toxicity forecasting



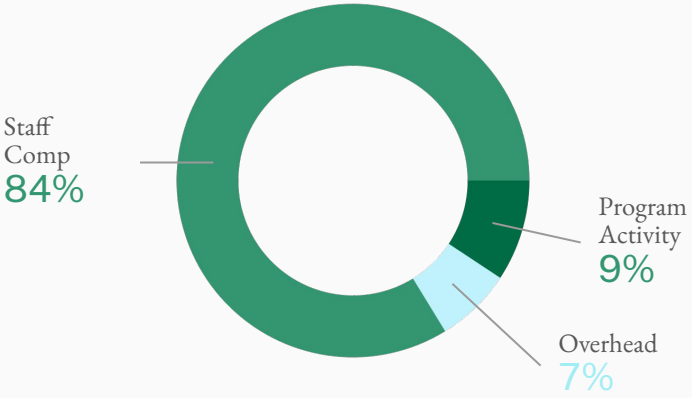
*See our annual and Q1 OKRs in the appendix

What We Need

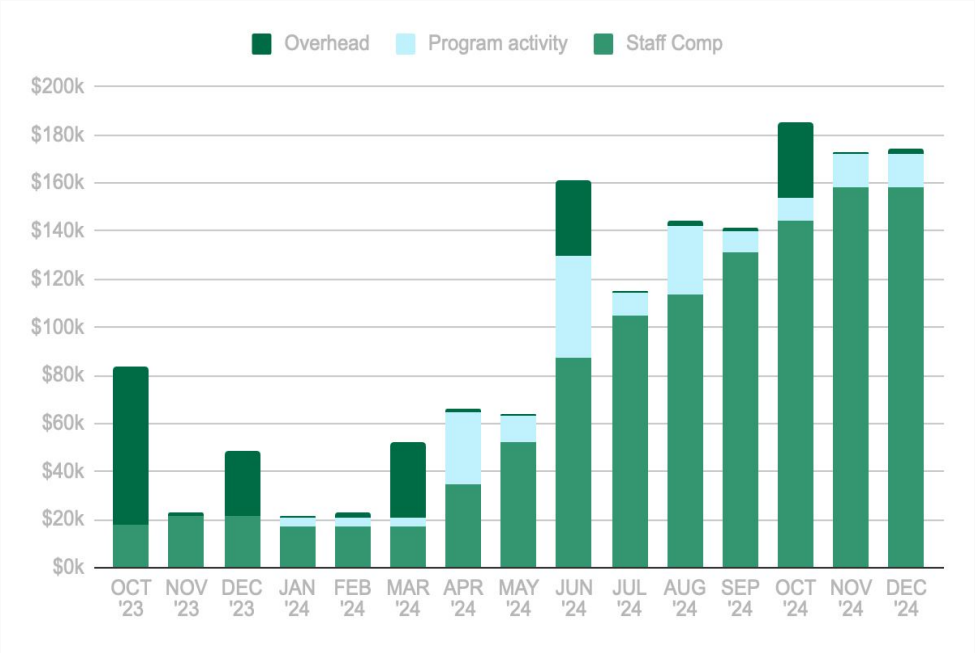
$2.5-3.5M Target

to focus on key results through end of 2025

Use of Funds



Estimated monthly spend



How Involved Would You Like to Be?

Atlas

	Financial Support ($k)	Target # of Supporters
→ **Help set and steer our goals** Our mission resonates with you and you want to be very engaged. • Nominate a board member • Receive invites to our private workshops • Help us set annual and quarterly plans • And everything from the next 2 tiers	500+	2
→ **Be a key supporter & stakeholder** You believe in us as one of a large portfolio of approaches. • Give feedback on our annual and quarterly plans • And everything from the next tier	100 - 500	5
→ **Help demonstrate broad appeal** Support us with your reputation and network • Be listed as a supporter on the website • Receive complimentary access to all our public events	25 - 100	10
→ **Collaborate on your priorities** Collaborate on your priorities: • Running a conference / workshops • Hosting interns • Hiring an internal comms/media team • Fast-tracking plans to a self-sustaining business model • Prioritize a use case (e.g. secure electrical grid)	(variable)	(variable)

Want to support us? Let’s chat!

https://calendly.com/miyazono/30-min

Appendix: Additional Planning Links

2024 Annual and quarterly OKRs

https://docs.google.com/spreadsheets/d/15fSq-c9_huPqhHJ5B3gpwGn0gcCYXxIgWKmaSRGxS6o

2024 Gantt Chart:

https://docs.google.com/spreadsheets/d/1dzfNB_C36NrSQF6gF70e7Vlb7ckm4ydzY4nMvmOr18A

Line-item budget forecast here (sorted by decreasing marginal value):

https://docs.google.com/spreadsheets/d/13TrwA6X8yOfLKoRPeqVeHdnRK9_td3MOFMtvEtlM9Hw

Update emails:

https://groups.google.com/a/atlascomputing.org/g/updates

Appendix: Potential Org Chart

Last updated Jan 29

Atlas

Base funding request	4
Ambitious funding request	3
Max funding request	3