### The Problem

<table>
<thead>
<tr>
<th>Domain</th>
<th>Benefit</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic</td>
<td>Improve human systems</td>
<td>Outpace human oversight</td>
</tr>
<tr>
<td>Cyber</td>
<td>→ Fix software vulnerabilities</td>
<td>! Make new computer viruses</td>
</tr>
<tr>
<td>Bio</td>
<td>→ Cure diseases</td>
<td>! Make new pathogens</td>
</tr>
<tr>
<td>Law</td>
<td>→ Patch legal ambiguities</td>
<td>! Find legal loopholes</td>
</tr>
<tr>
<td>Society</td>
<td>→ Reduce information overflow</td>
<td>! Create disinformation</td>
</tr>
</tbody>
</table>
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.
The Solution

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**

User(s) → Informal Spec → Solution Generator → Candidate Solution → Approved Solution

**But this**

User(s) → Specification Assistant → Solution Spec → Solution & Proof Generator → Candidate Solution & Proof → Solution Checker → Verified Solution

Enable **scalable human review**
The Goal

An international standards organization for verifiably governable AI

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

Help ✅ each of these in high-leverage ways
(e.g. match great people to funding, prototypes, customers, and investments)

<table>
<thead>
<tr>
<th>Stage</th>
<th>Cyber</th>
<th>Bio</th>
<th>Law</th>
<th>Society</th>
<th>etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant experts aware of AI risk?</td>
<td>✅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification language exists?</td>
<td>✅</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI enhanced spec generators exist?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI enhanced solution generators exist?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI enhanced solution checkers exists?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Atlas

Done
We’re building
Others’ progress
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:

   → 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.

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

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
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.
Advisors

→ 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.
Why A Non-Profit

- 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...

Solution Generator

Candidate Solution

Solution Spec

Solution Generator

Specification Language

User(s)

Review

Natural Language
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](#).

**User(s)**

I'm trying to update the following FORTRAN code that does x

**Natural Language**

Formal specification languages for software already exist (e.g. Coq, Dafny, Lean), but take years to learn

**Specification Language**

Here's a specification in Lean for the code

**Specification Generator**

Here's code and a proof it won’t violate the spec

**Solution Spec**

**Solution Generator**

**Candidate Solution**

**Solution Checker**

**Review**

Cyber Example Product: Update Legacy Software
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.

Small Molecule → AlphaFold analog → Bioactivity Forecast Specification Language → Toxicity Forecaster → LD 50

All predicted interactions with different metabolic pathways

User(s) → Natural Language → Specification Generator → Solution Spec → Solution Generator → Candidate Solution → Solution Checker

→ Review

Design a drug to modify pathway x without causing harmful effects

Here’s a specification that formalizes ‘modify’ and ‘harmful’

Here’s a molecule and evidence it doesn’t violate the formalization of ‘harmful’
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

$925k Raised

Use of Funds

Staff Comp 84%
Program Activity 9%
Overhead 7%

Estimated monthly spend
# How Involved Would You Like to Be?

<table>
<thead>
<tr>
<th>Financial Support ($k)</th>
<th>Target # of Supporters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Help set and steer our goals</td>
<td>Be a key supporter &amp; stakeholder</td>
</tr>
<tr>
<td>Our mission resonates with you and you want to be very engaged.</td>
<td>You believe in us as one of a large portfolio of approaches.</td>
</tr>
<tr>
<td>500+</td>
<td>100 - 500</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

### Help set and steer our goals
- Nominate a board member
- Receive invites to our private workshops
- Help us set annual and quarterly plans
- And everything from the next 2 tiers

### Be a key supporter & stakeholder
- Give feedback on our annual and quarterly plans
- And everything from the next tier

### Help demonstrate broad appeal
- Be listed as a supporter on the website
- Receive complimentary access to all our public events

### 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)
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/1fSq-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

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

CEO
Evan Miyazono

COO
(to hire)

CTO
(promote a lead; not new headcount)

Product Lead
(to hire)

FV+AI Research Lead
(to hire)

Event Lead
(to hire)

Bio Spec Lead
(to hire)

Talent Lead
(to hire)

Software Lead
Daniel Windham

AI Developer
(to hire)

FV Application Engineer
(to hire)