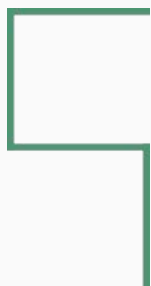


Atlas Computing



AI capabilities are outpacing human review

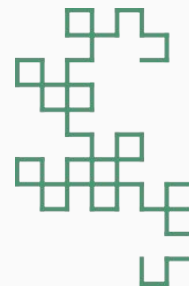
TODAY:
Simple requests to AI



Simple responses
Review is EASY



SOON:
Complex requests

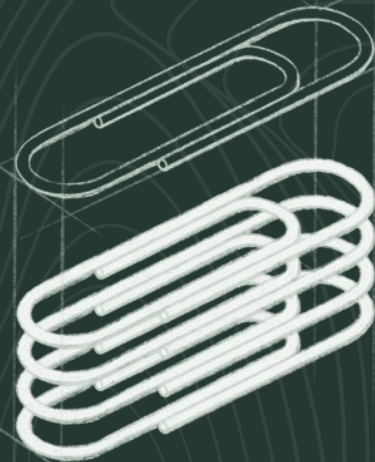


Complex responses
Review is... SKIPPED?

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 from using “**profit as proxy for good**” could destabilize any human system.



AI makes disruption easier than stabilization

Domain

Systemic



Cyber



Bio



Law



Society

Benefit

Improve human review

→ Fix software vulnerabilities

→ Cure diseases

→ Patch legal ambiguities

→ Reduce information overflow

Risk

Outpace human oversight

! Make new computer viruses

! Make new pathogens

! Find legal loopholes

! Create disinformation



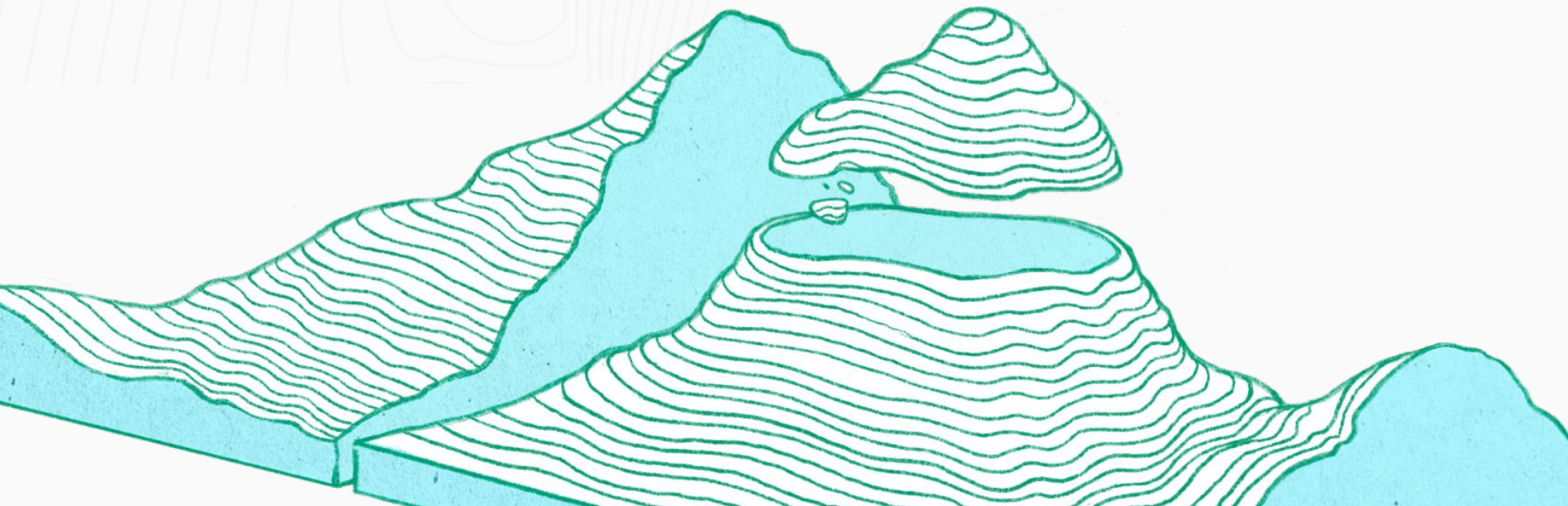
Everyone wants
human-level AI agents.

No one can define “good behavior for humans.”

“Good AI” is...

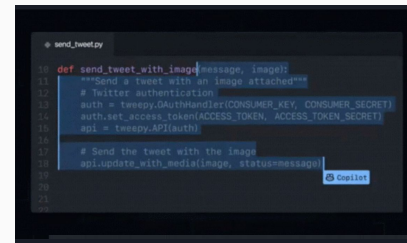
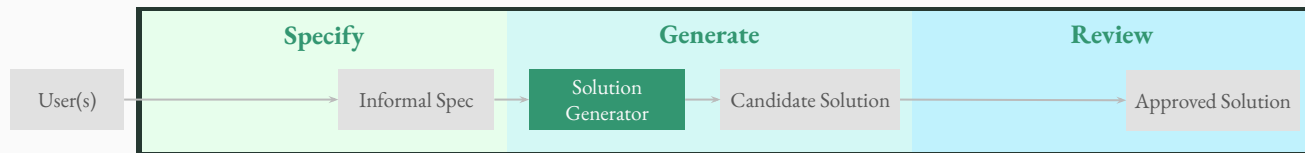
- harder,
- subjective,
- a moving target

Make tools
that the AI proves it's obeying.

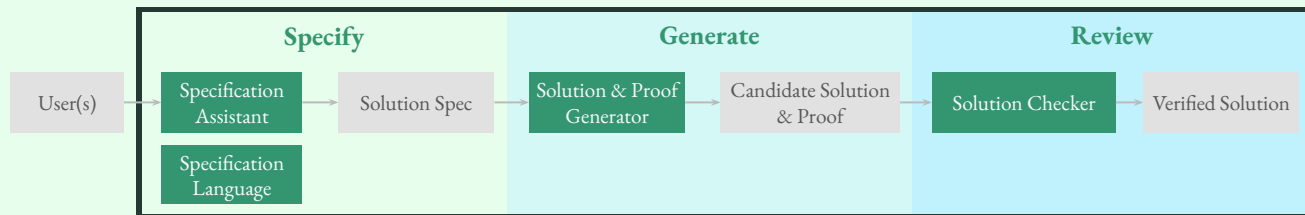


Specifications: human-reviewable intermediate outputs from AI

Not this



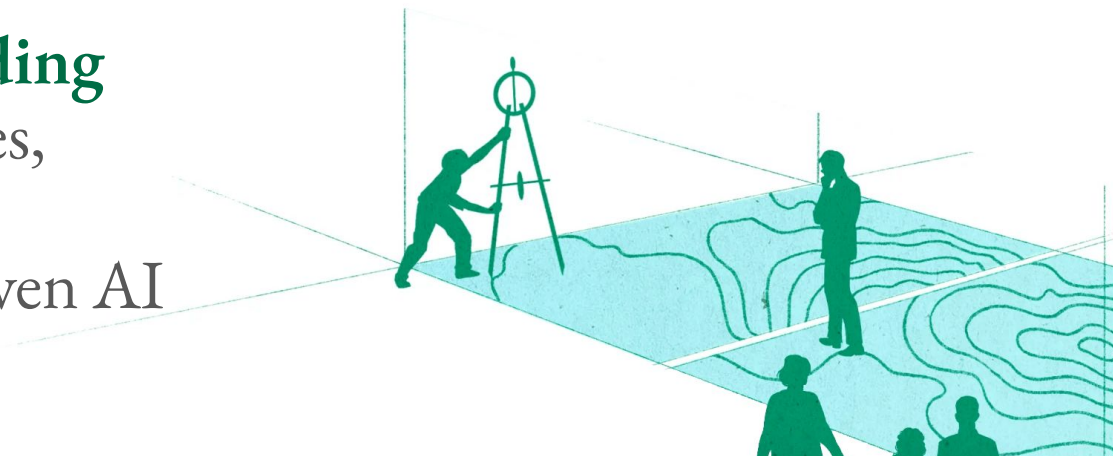
But this



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






This architecture is the best way to govern AI,
but it only works if everyone uses it.












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



Build and deploy tools one vertical at a time

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

-  Done
-  We're building
-  Others' progress

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



Domain 1

Cyber

- **Goal**
Leverage growing AI capabilities to scale the use of existing spec languages for software
- **First Deliverables**
Formal Methods + AI 2-pager and Coq to Lean translation
- **Current step**
Project list of 15 tools to upgrade & formally verify legacy software



Domain 2

Bio

- **Goal**
Make a new spec language
- **Current Step**
Refining proposal for a toxicity forecasting competition and recruiting a project lead



Building a

Community

- **Co-organized 4 workshops**
See our events page
- **Organized an email list**
Organizing discussions among collaborators

AI risk is better averted one domain at a time.

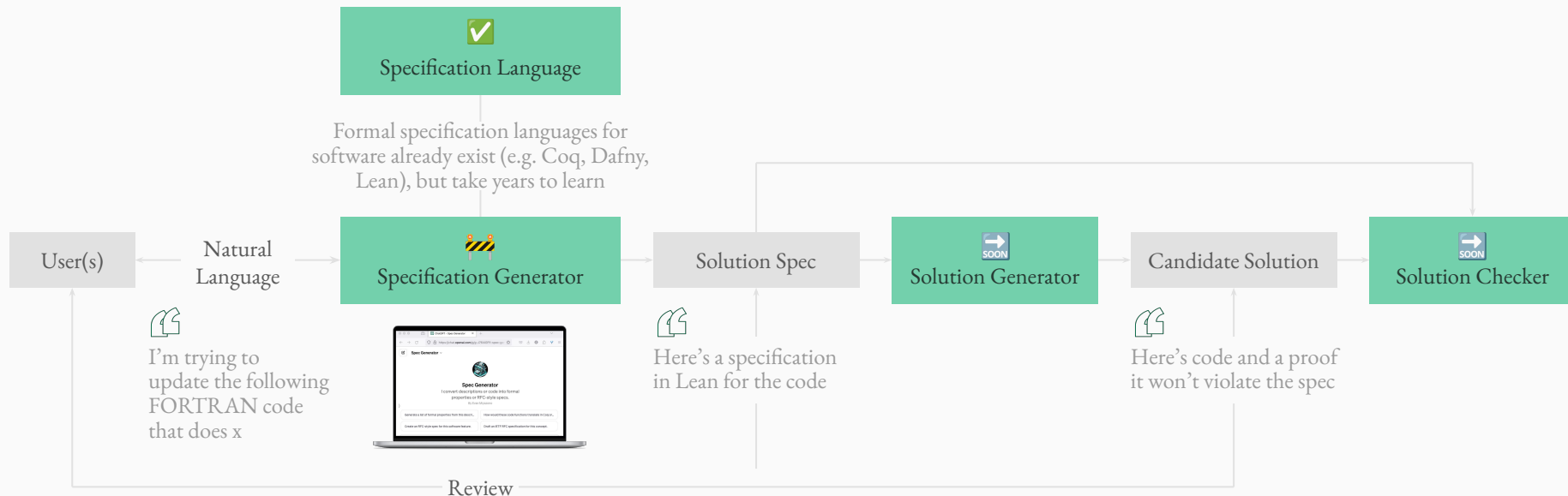
Rather than attempting to solve the whole problem at once.

LLMs can scale Formal Verification.

(an existing but costly specification language for software)

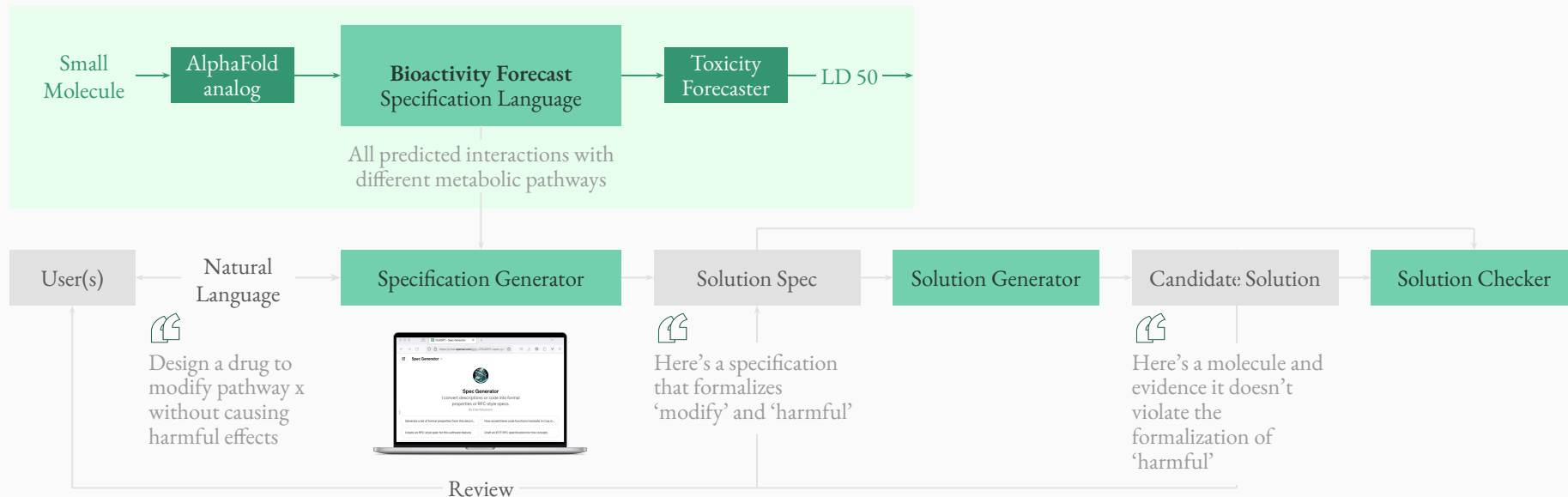
Updating software is critical but risks breaking capabilities

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



The Team

Atlas



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.

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

→ **Produce** persuasive evidence

that LLMs are ready to scale formal verification in real-world systems

→ **Identify** a stakeholder

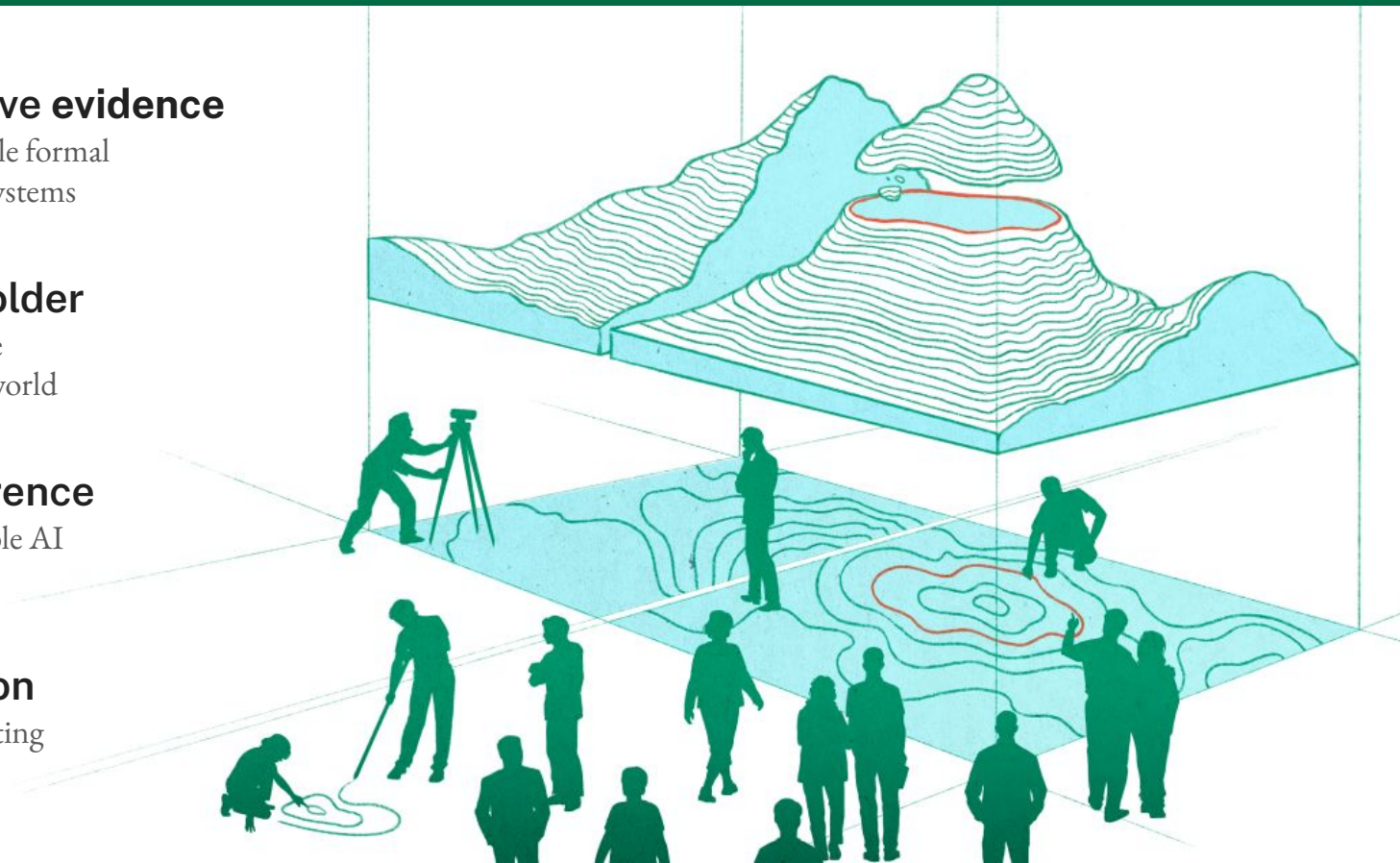
who can use LLMs to scale formal verification in the world

→ **Organize** a conference

for 100+ people on provable AI safety properties

→ **Start** a competition

to advance toxicity forecasting

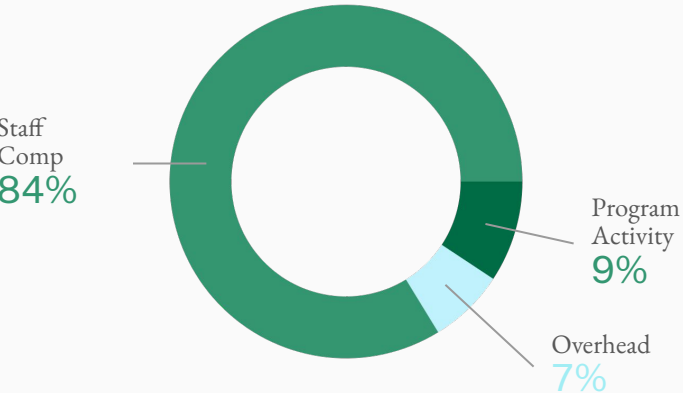


\$2.5-3.5M Target

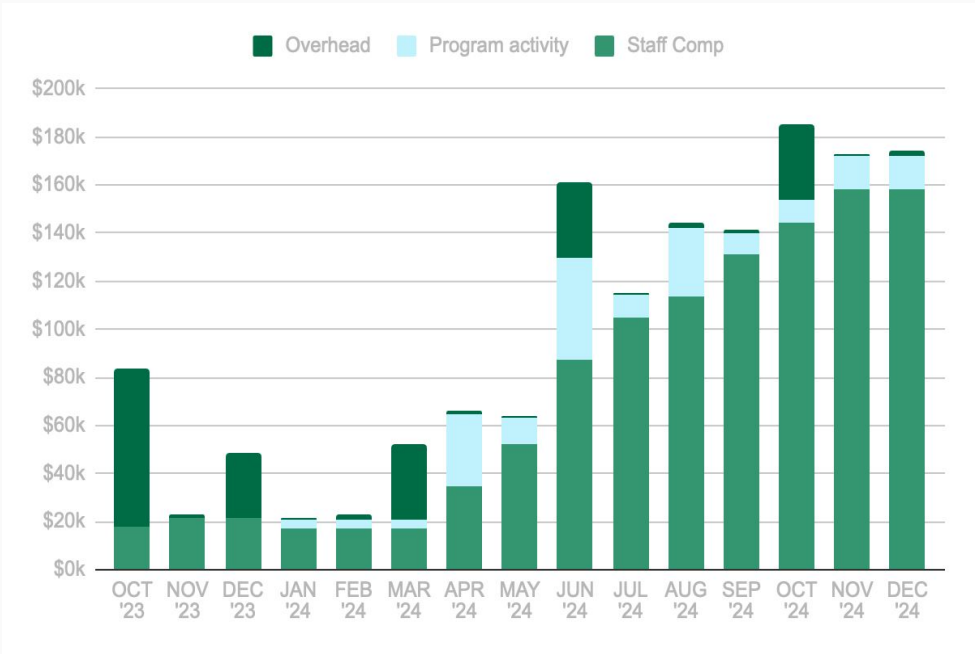
to focus on key results through end of 2025



Use of Funds



Estimated monthly spend



Thanks for reading

Atlas

Want to support us? Let's chat!

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



Appendix: Additional Planning Links

Atlas

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:

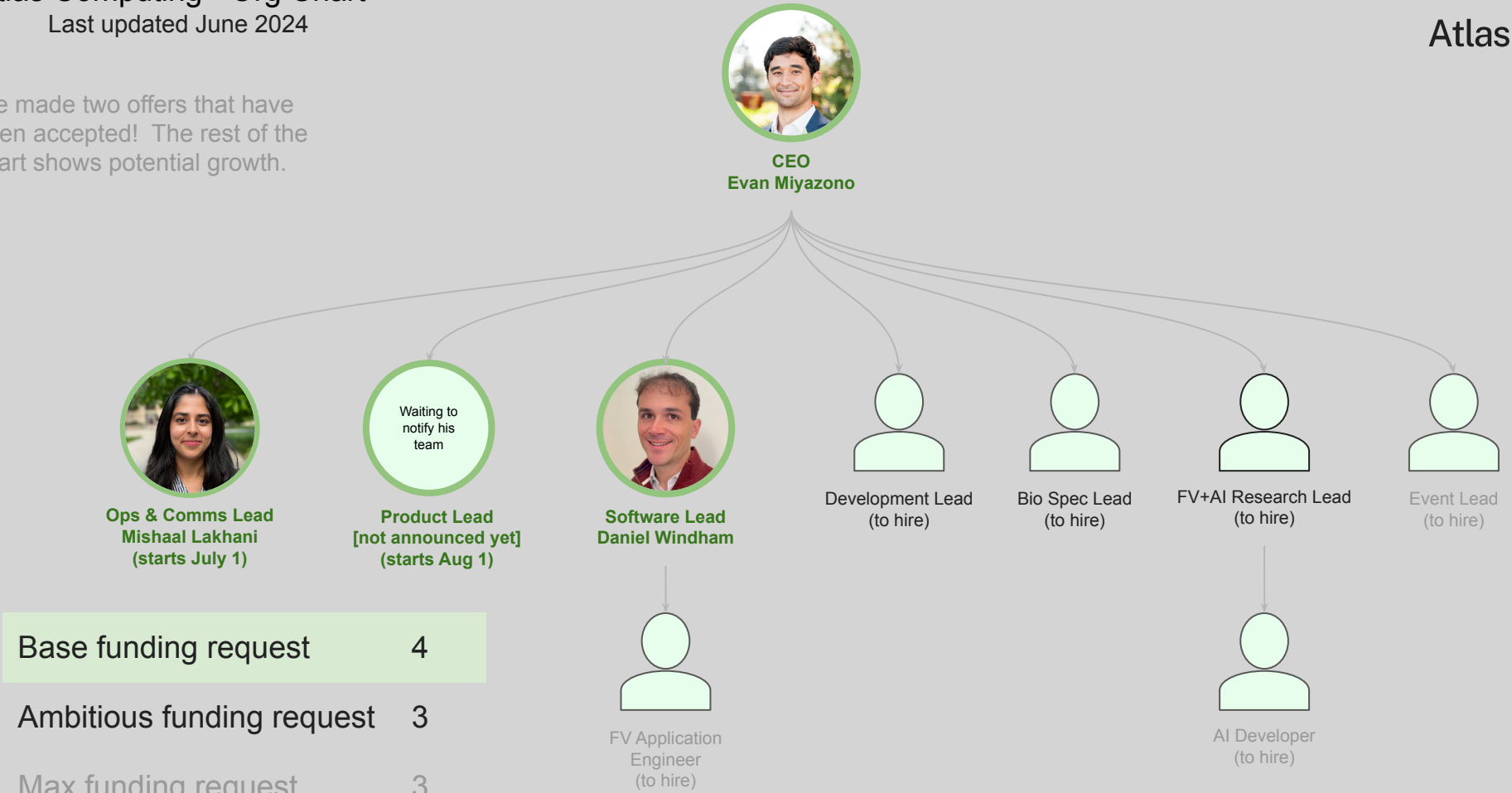
<https://groups.google.com/a/atlascomputing.org/g/updates>

Atlas Computing - Org Chart

Last updated June 2024

Atlas

We made two offers that have been accepted! The rest of the chart shows potential growth.



Base funding request 4

Ambitious funding request 3

Max funding request 3