Small AI Teams with Huge Impact — Vikas Paruchuri, Datalab

371 views Jul 15, 2025 AIEWF 2025 Complete Playlist

We scaled Datalab 5x this year - to 7-figure ARR, with customers that include tier 1 AI labs. We train custom models for document intelligence (OCR, layout), with popular repos surya and marker.

I'll talk about a new approach to building AI teams, including lessons I learned from Jeremy Howard, and how we manage building popular repos, scaling revenue, and training models with a tiny team.

About Vikas Paruchuri
CEO of Datalab

Datalab: 40k stars, 7-figure ARR, SoTA models, team of 3

Vikas Paruchuri

VikParuchuri vik@datalab.to vikparuchuri

Hi, I'm Vikas. I spent the last year:

1. **Training SoTA models for document intelligence**
2. **Building repositories with 40k+ GitHub stars**
3. **Leaving my AI research job**
4. **Starting a company and raising a seed round**

Introducing Datalab

Team

Made our first hire in January, now a team of 4 (myself, Tarun, Sandy, Faraz)

Revenue

Grown revenue 5x since January, at 7-figure ARR

Customers

Customers include tier 1 AI labs, universities, F500, AI startups

Vik Paruchuri

Sandy Kwon

Tarun Ram Menta

Today's Focus

Our Approach

How we've grown with a small team

Team Philosophy

My approach to building high-performing teams

Scaling Differently

> Headcount ≠ productivity

Lessons from Dataquest

1 **Growth Phase**

Scaled to 30 people and $4M in ARR during COVID

2 **First Reduction**

Post-COVID layoffs: 30 → 15 people

3 **Second Reduction**

15 → 7 people

4 **Productivity increase**

Productivity and happiness increased ~2 months after each reduction

Why Productivity Increased

Specialist Limitations

Specialists couldn't flex outside their areas. Often not solving key problems

Meeting Overload

Multiple teams and middle management required a lot of syncing

Remote Challenges

Remote culture required heavy process to maintain alignment

Experience Mix

Senior people got tied up in management, code review, etc

The Golden Period

Initial Golden Period

Everyone aligned and building quickly - build the core thing that drives revenue

Expansion Hiring

Hiring to fill edges around the core

Bureaucracy

Lots of syncs, meetings, unclear priorities

This pattern is common in startups. Think search with Google or Windows with Microsoft. What if that golden period could last forever?

Jeremy Howard's Philosophy

While working with Jeremy at answer.ai, I learned about his approach to preserving the golden period.

Small Team

Hire <= 15 generalists who build

AI Augmentation

Fill in the edges with AI and internal tooling

Simple Tech

Use boring, reliable technology

Cultural Requirements

Full-Stack Understanding

Team members need to understand and build across every aspect of the company

High Trust

A collaborative environment with some top-down guidance is essential

Customer Focus

Everyone must understand the customers and their needs deeply

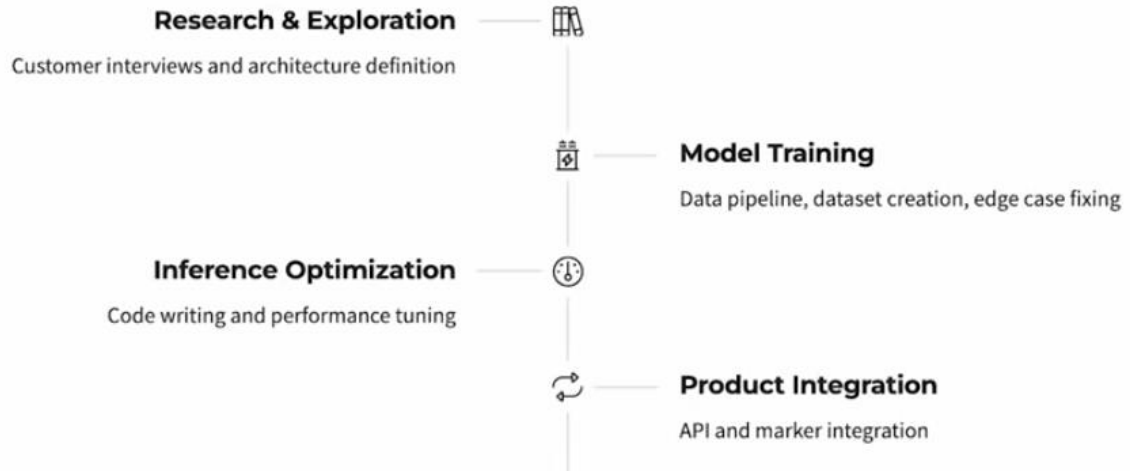
Surya OCR 3 Example

500M	90+	99%
Parameters	**Languages**	**Accuracy**
Model size optimized for performance	Comprehensive language support	On challenging internal benchmarks that include math

Our recently shipped (not yet announced) OCR model is fast, supports every major language, handles block and inline math, does character-level bounding boxes, and uses PDF text as grounding.

Cross-Functional Development

Tarun (Research Engineer) and I handled the entire process from research to product integration. We were able to use AI to focus on the highest leverage parts of each group.



Benefits of Our Approach

Traditional Scaling

Hire individual teams to own each piece

- Lossy interfaces between teams
- Inefficient comms and syncs
- Slow feedback loops

Our Approach

Generalists with AI augmentation

- Seamless context sharing
- Tight product/model integration
- Rapid feedback cycles

With AI, we focus on high-value work while automating the rest.

More People ≠ More Productivity

Making This Model Work: Cultural

Hire Senior Generalists

Senior ≠ years of experience!

Avoid Overcomplication

Do you really need that kubernetes cluster?

Work In-Person

Avoid heavy process and remote syncs.

Making this model work: Architectural

Re-use components

Across services (on-prem, API)

Keep technology simple

Server-rendered html, clean API, etc

Clean, modular code

That is easy for AI to add to

Keep Everything Simple!

Code

Clean, readable, and maintainable by both humans and AI

Architecture

Fewer moving parts, use managed tools where possible

Process

Minimal bureaucracy, high trust, continuous discussions

Scaling: models fill in edges

Problem

Every customer has a slightly different way that they want to parse docs

Traditional solution

Use forward deployed engineers and a consultative model

New solution

Train your model to handle the complexity (agent loops, etc)

When does this model fail?

> **We're seeing companies scale to big ARR with small teams - Gamma with 30 people and $50M ARR is one example.**

> **The key is being able to say no, and having the discipline to separate the notion of increasing productivity from increasing people.**

> **Still unknown if this can work forever, but all of the edges and extra complexity you take on are choices.**

Our Three Core Roles

We only hire for three job titles with mostly overlapping responsibilities: Research Engineer, Fullstack Engineer, and GTM.

Research Engineer

- Talk to customers
- Define model architectures
- Build datasets
- Train models
- Write inference code
- Build API
- Build product

1

2

Fullstack Engineer

- Talk to customers
- Build datasets
- Write inference code
- Build API
- Build product

Go To Market

- Talk to customers
- Product positioning
- Sell to customers
- Build product

3

[redacted]ee roles share customer interaction, while technical responsibilities overlap between Research and Fullstack Engineers.

Hiring the Right People

Politics are the death of small teams. We want people who only care about the work, team, and customers.

Competitive Compensation

Pay top of market salary

Meaningful Work

Big challenges and scope

Cultural Fit

Screen for low ego and GSD

Patience

Hire very slowly to find the right people - pressure to fill roles means suboptimal tradeoffs

Scaling Productivity, Not Headcount

There's pressure to scale headcount as you raise money, but we focus on scaling productivity instead.

Raise salary bands

Attract and retain top talent

Compute Investment

More spending on compute to accelerate training, unlock research

AI Tools

Invest in AI tools that multiply productivity

Join Our Team

If it sounds exciting to work in an environment like this, shoot me a message, ideally with something impressive you've built. We're hiring for RE, fullstack engineer, and GTM.

VikParuchuri

vik@datalab.to

vikparuchuri