



scale

Data: The Raw Material Fueling AI

Vijay Karunamurthy, Field Chief Technology Officer

The 3 Pillars of AI

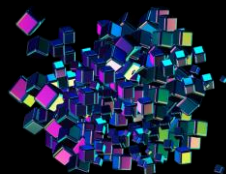
COMPUTE

GPU Chips



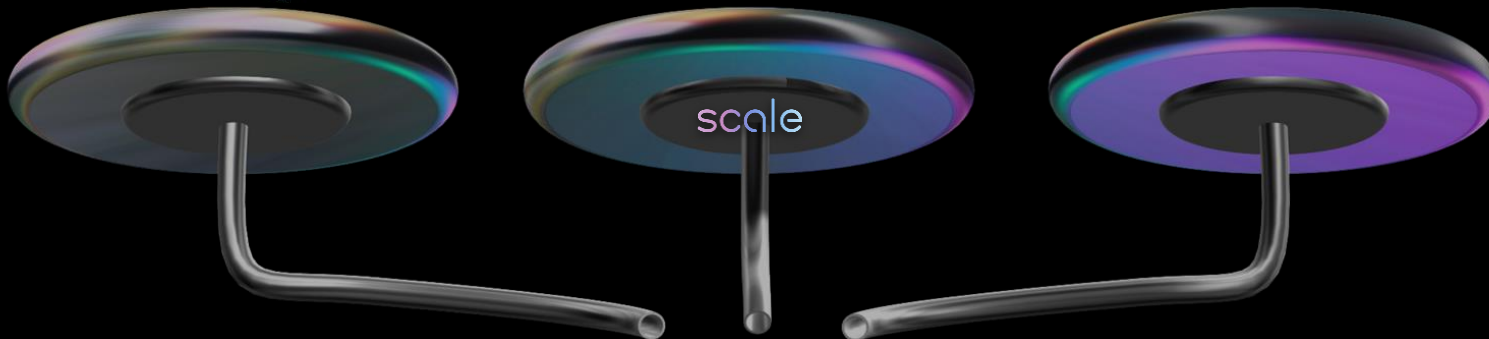
DATA

Training Data + Eval



ALGORITHMS

Talent



AI Advancement



Scale has powered every major AI breakthrough across computer vision and generative AI



AUTONOMOUS VEHICLES

2016

COMPUTER VISION



U.S. Department of Defense

2020

U.S. GOVERNMENT

∞ Meta



2021+

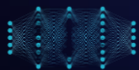
GENERATIVE AI



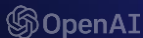
Accelerating the development of AI applications...



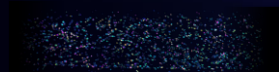
Chips



Models



ANTHROPIC



Data

scale

OUR MISSION

Accelerate the development of AI applications

OUR INSIGHT

These models *are* their data. The only path to improvement is through quality data

OUR WORK

- Dramatic improvements to model accuracy through advanced training techniques
- Help model to run real-world ready agent work flows
- Enable next-gen models to do X domain in Y language



...for Builders, Enterprises, and Governments

Frontier Model Builder

Scale evaluates and generates more GenAI data per day than any company on the planet

Leading Enterprise

Enterprises use Scale's platform to get GenAI apps into production faster through better data

Governments

Scale powers major AI U.S. government programs with the leading commercial technology



U.S. AIR FORCE



U.S. ARMY

ADEPT





Pioneer of model customization, fine-tuning, and evaluation practices



Scale is OpenAI's preferred fine-tuning partner for GPT-3.5

Scale is Meta's launch partner for enterprise Llama 3.1 405B

Scale GenAI Platform is now available on the Microsoft Azure Marketplace

Scale becomes the first AWS model customisation and evaluation partner on AWS Marketplace

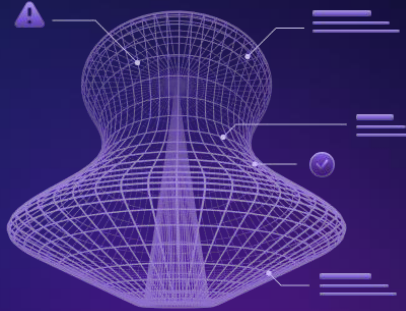


Scale builds custom full-stack GenAI applications for governments and enterprise



National LLM

Full development and maintenance of a Foundational LLM tailored to language, cultural context, and history



Custom AI Use Cases

End-to-end development for Custom AI applications in enterprise or in government



Model Enhancement

Evaluations and high quality data for significant improvements in model performance and safety



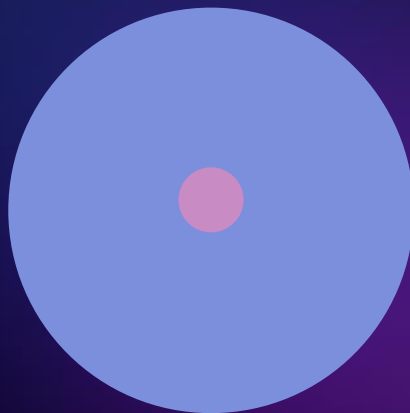
Proprietary data and experts are the most valuable assets in GenAI Systems

Every organisation is going to need a scalable platform and data engine that uses their proprietary data to build specialised GenAI models

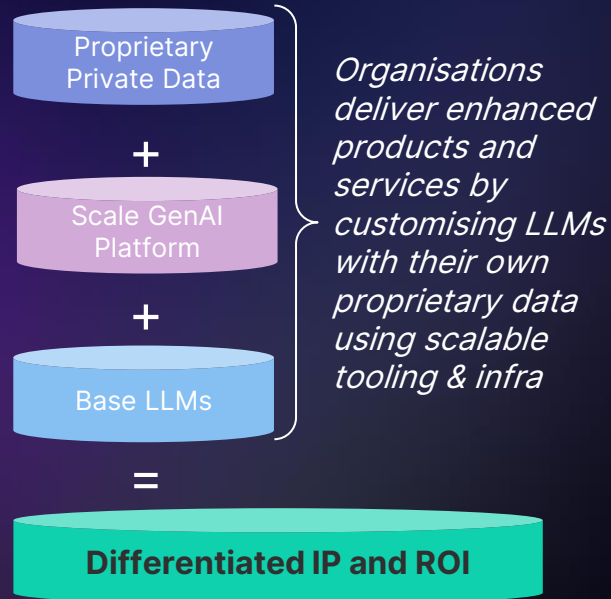
- Public data
- Private data



Current model data mix

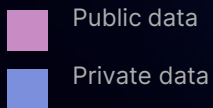
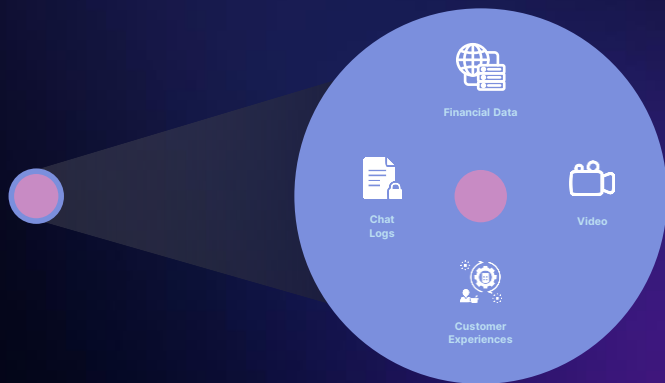


Total available data sources





Enterprise & Government data edge must be leveraged in model output



Current LLM data mix

Total available data sources

Enterprise data = 10x of public data

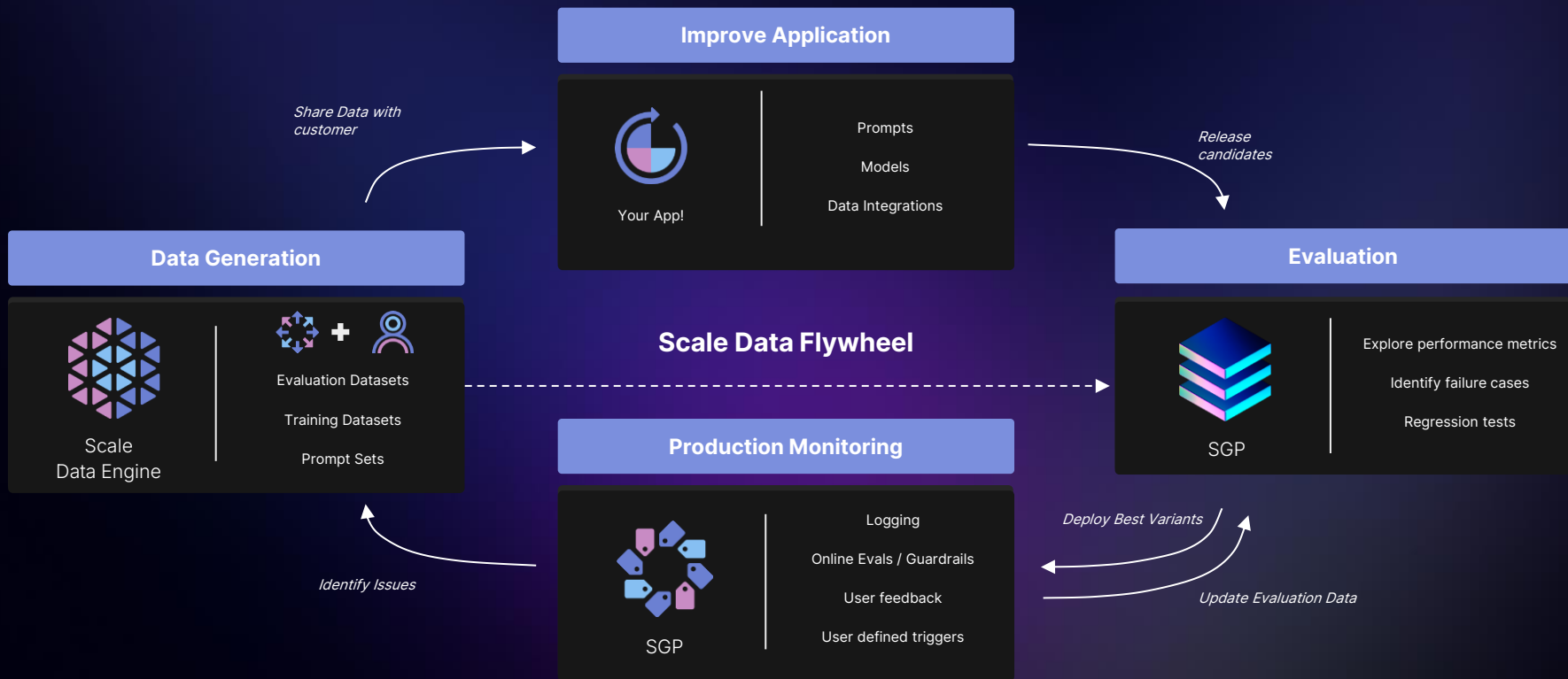
Yet models today are powerful *generalists*, but not the *experts* that governments and enterprises need.

The Expertise Problem

Models will only be “taught” to become *experts* by curating the right - *relevant* and *structured* - information as input.



Scale Generative AI Platform Data Flywheel





Leaders in AI test and evaluation



Safety, Evaluation, and
Alignment Lab



United States
Department of Defense,
Chief Digital and Artificial
Intelligence Office



Custom LLMs in production need a “Trust Feedback Loop” for scaled deployment

- ✔ Trusted and safe GenAI
- ✔ Production-grade performance
- ✔ Faster development cycles
- ✔ Lower TCO through OSS customization
- ✔ Full visibility and control



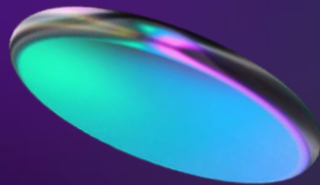
Build

Architect your GenAI Applications



Evaluate

Identify performance gaps



Improve

Generate Data (Training & Evals)

Fine-tune Gen AI Applications

ITERATIVE & CONTINUOUS
DEVELOPMENT



Leveraging proprietary data for custom AI applications is the future



Data Analysis



Customer Service



Process Automation



Content Generation



Knowledge Management



Multilingual Support

The image features three glowing, metallic discs floating in a dark, gradient background. One disc is positioned in the upper left, another in the lower left, and a third in the lower right. Each disc has a bright, multi-colored rim that transitions from purple to blue to green. The word "scale" is centered in the middle of the image in a light blue, sans-serif font.

scale