

Harvest AI

Backed by  Combinator



Samarth Kadaba

EE @ Stanford
Quant @ SMBC
Researcher @ SAIL



Josh Sanyal

CS @ Stanford
SWE @ Meta
Quant @ Jane Street



Karan Bhasin

CS @ Stanford
SWE @ Illumio
SWE @ Dapper Labs



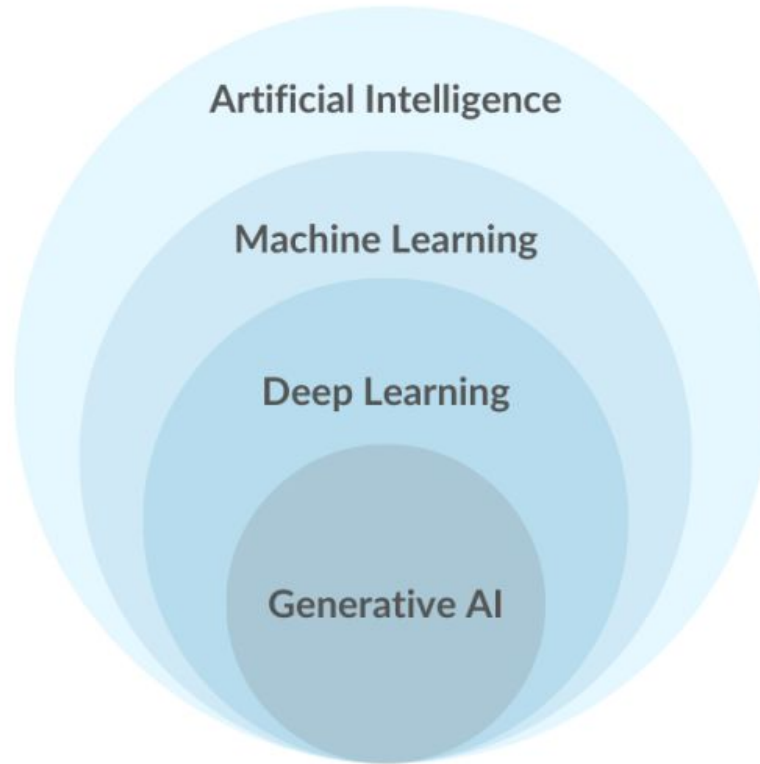
Avi Basnet

CS @ Stanford
SDE @ Amazon
SWE @ NASA

Experience building:

- Privacy aware chatbots (ChatGPT), OpenSource LLMs on internal servers and closed networks
- Diffusion/continuous generative models for anomaly detection
- Multivariate and nonlinear models for trade data (incorporating structured and unstructured data)
- Database migration to AI-compatible vector databases (efficient for natural language search and document drafting)
- Clinical notes documentation and drafting
- Published computational biology research in Nature

Understanding data distributions with Generative AI



Understanding data distributions with Generative AI

Language

- Draft documents
- Respond to queries using data
- Search databases with natural language (information retrieval)
- Time series analysis

Multimodal and Heterogenous data

- Synthetic multimodal data generation
- Probabilities for high-dimensional variables
- Anomaly detection
- Process modeling (transforming distributions)

Utility across the value chain

**60-70% reduction in
process timelines**

**> 30% reduction in
operational costs**

**> 40% reduction in
project delivery timelines**

Use cases by category

1) Summarize/Create

- Summarize external/internal research for scientists
- Summarize clinical findings from data sources
- Automatic drafting of SOPs, CAPAs, deviation reports, regulatory filings and more
- Clinical trial design and documentation based on historical/similar studies

2) Search

- Search over clinical data to compare cohorts and outcomes only from documentation (avoid I/E training)
- Find relevant external/internal research with natural language

3) Insights

- Process testing with synthetic MFG data (digital twin)
 - Parameter optimization
- Batch/lot forecasting
- QA insights from time-series reporting (deviations, etc.)

Use Case #1: Predict batch quality using APH characteristics

- High variability/heterogeneity in incoming APH materials
- Predict batch quality and manufacturability using APH characteristics
 - Identify predictive attributes of MFG outcomes
 - Re-prioritize, re-schedule batches based on APH and in-process observations
 - Predict < or > day 7 harvest
- Generate synthetic data to build robust models from little real-world data

Use Case #2: Deviation management and reporting

- Deviations inhibit on-time delivery of batches both in manufacturing and clinical trials
 - Many deviations are repeat (human error, machine failures, etc.)
- Building robust deviation tracking and management early reduces reporting effort as manufacturing scales
 - Identify root causes and trends in deviations
 - Automate CAPA and deviation reporting (first drafting saves labor hours)
 - Efficiently search over and find relevant CAPAs/deviations from the past

Use Case #3: Iterative scenario modeling: clinical trial design

- Leverage historical external/internal data to build predictive models for trial outcomes (ex. regulatory approval)
- Identify trial parameters to maximize specific objective function
 - total enrollment
 - statistical power
 - Cost-effectiveness
- Hypothetical trial analysis from scarce data sources