

---

# PRANAV KUMAR SINGH

Fremont, CA 94560 | 716.445.6035 | [www.linkedin.com/in/pranavsingh](http://www.linkedin.com/in/pranavsingh)

---

## Vice President of Engineering Engineering Leader | Generative AI | Multi-Agent Systems | Scalable AI Platforms

Visionary technology executive with 18+ years of experience leading high-impact engineering teams, architecting and delivering enterprise-grade AI platforms, and spearheading innovation in GenAI, large language models, and multi-modal agents. Proven success in designing and scaling developer-facing AI products, from early concepts to production deployment. Patented inventor and thought leader in conversational AI, orchestration, and high-performance systems serving global enterprises.

---

### PROFESSIONAL EXPERIENCE

---

**SOUNDHOUND AI Inc**, Santa Clara, CA (NASDAQ: SOUN)  
*Vice President of Engineering, Aug 2014 - present*

2014 - present

- Founder and lead developer of [CaiNET - SoundHound's core Machine Learning technology](#) for Natural Language Understanding, significantly improving accuracy and contextual understanding in conversational AI.
- Founded and led development of the Agentic AI architecture, creating a robust multi-agent framework now powering cross-vertical solutions in automotive, QSRs, food ordering, customer service and enterprise domains.
- Directed engineering efforts for LLM fine-tuning, distillation, and function-calling frameworks, enabling hybrid AI systems with deterministic and generative capabilities.
- Spearheaded the development of modular engineering frameworks and reusable AI components, dramatically accelerating time-to-market for new product lines.
- Contributed to the development of internal foundational ASR models, supporting multiple domains and improving recognition accuracy for specialized verticals.
- Collaborated with product, research, and field engineering teams to deliver scalable AI-powered solutions, ensuring high technical quality, reliability, and scalability.
- Led the design and deployment of real-time data pipelines and ML infrastructure (Kafka, Spark, Docker, Kubernetes), ensuring high-availability and performance across cloud and edge environments.
- Built and scaled innovations, including [Employee Assist](#), a modular platform using internally trained LLMs to empower customer-facing teams, and [Vehicle Intelligence](#), enabling voice interfaces for automotive partners like Lucid, Hyundai, and Stellantis.
- Served as SoundHound's technical evangelist at industry conferences including CES and NVIDIA GTC, showcasing AI innovation and driving awareness of platform capabilities.
- Scaled cross-functional teams across data, ML, and product engineering, Instrumental in scaling the engineering org from 30 when I joined, to 600+ employees and led it through IPO (Nasdaq: SOUN, \$3B+ valuation).
- **Patents (7):** solidifying the company's position as a leader in conversational AI.
  - Bidirectional probabilistic natural language rewriting and selection (Issued Mar 24, 2020 US 15/726,394)
  - Predicting human behavior by machine learning of natural language interpretations (Issued May 24, 2019 US 15/425,09)
  - Natural Language Module Store (Issued Oct 16, 2018 US 14/954,810)
  - System & Method for Detection & Correction of a Query (Issued Mar 1, 2022 US 16/561,020)
  - Machine learning system for digital assistants (Issued Aug 20, 2024 US 17/350,294)
  - Token confidence scores for automatic speech recognition (Issued Feb 11, 2025 US 17/649,810)
  - Neural sentence generator for virtual assistants (Filed Nov 19, 2021 US 17/455,727)

**Arjun Aarya Ventures, (<https://www.arjunarya.com/>)**

2022 - present

*Founder, Jan 2022 - present*

- Conceived, designed, and built consumer-grade mobile products under the Wanna-Be-So™ brand, targeting productivity and lifestyle segments with complete focus on UX and privacy.
- Developed Wanna-Be-So™ Organized, a task planner with intelligent categorization and recurring task management – ranked for simplicity and clarity.
  - ios: <https://tinyurl.com/wannabesoios>
  - android: <https://tinyurl.com/wannabesoandroid>
- Created Wanna-Be-So™ Mindful, a personal memory assistant allowing users to store and retrieve multimedia memories locally and securely.
  - iOS: <https://tinyurl.com/mindfulios>
  - android: <https://tinyurl.com/mindfulandroid>
- Personally did all aspects of product engineering – from frontend Flutter UI to backend local storage logic and deployment to iOS and Android stores.
- Applied deep user empathy and lean methodology to test, iterate, and ship compelling features that solve real-world pain points.

**ORACLE AMERICA Inc, Redwood City, CA**

2011 - 2014

*Senior Software Engineer/ Technical Lead, Jan 2011 - Aug 2014*

- Led the end-to-end design and development of a Big Data-based unstructured content enrichment platform to improve enterprise search results for Oracle's Endeca engine.
- Founded and architected the "Enrichment Engine" framework, a modular system to apply NLP and ML-based transformations (e.g., sentiment analysis, named entity extraction, wikification) to large-scale datasets in HDFS using Spark, Hive, and Pig.
- Created a REST-based UI platform to host, test, and demo enrichment plugins – empowering internal sales, presales, and product teams to showcase the product to clients with intuitive workflows.
- Worked closely with Product Management and Customer Engineering to turn early prototypes into reliable tools used across enterprise accounts.
- Improved GTM velocity by enabling self-serve enrichment workflows for internal stakeholders and customers – shortening feedback cycles from weeks to days.
- Spearheaded Search Engine as a Service (SEaaS) and Search Federation as a Service (SFaaS) efforts – turning search into a plug-and-play cloud offering federated across products like Taleo and RightNow.
- Managed sprint planning, backlog grooming, and Agile delivery as Scrum Master, enabling fast iteration and tight alignment with customer needs.

**WIZE COMMERCE (Formerly NEXTAG Inc), San Mateo, CA**

2008 - 2011

*Software Engineer II - Search backend, Dec 2008 - Jan 2011*

- Developed and optimized backend infrastructure for real-time billing, clickstream analytics, and search behavior tracking, enabling deeper GTM insights and campaign targeting.
- Designed a Spider Management Framework to intelligently detect and filter bots – preserving system integrity and improving data quality for customer analytics and reporting.
- Built internal tools to visualize requester behavior and perform root-cause analysis of anomalous traffic – empowering marketing and revenue teams to take proactive actions.
- Enabled front-end experimentation through A/B test integration, allowing product and GTM teams to measure real-world impact of new user-facing features on search and conversion.
- Spearheaded image server optimizations, improving SLA performance and site responsiveness, directly impacting user engagement and ad revenue.
- Delivered GTM-aligned innovations by working across engineering, analytics, and performance teams to surface meaningful customer and operational insights.

**PAYPAL (EBAY), San Jose, CA**

2008

*Software Engineer, July 2008 – Nov 2008*

- Developed the Dynamic Navigation platform to enable product managers to update PayPal's navigation experience dynamically, increasing agility for GTM teams responding to business needs.
- Contributed to the Reputation Number system, designed to surface user engagement metrics – early experimentation with gamification and credit systems to enhance user behavior visibility.

**IBM, Bangalore, India**

2005 – 2007

*Software Engineer*

- Played a core engineering role in building middleware infrastructure (COS) for Sprint Communications on IBM's telco delivery platform using Java and EJBs.
- Led delivery of full release cycles and mentored new engineers, collaborating with global teams to ensure timely and scalable rollout for enterprise clients.

---

## EDUCATION

---

**STANFORD UNIVERSITY Graduate School of Business, Stanford, CA**

2018-2019

**Executive Education: Stanford Ignite - GSB program for Innovation and Entrepreneurship**

**Credit Course in Entrepreneurship, Global Entrepreneurial Marketing, Spring Semester, 2013**

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY, Cambridge, MA**

2014

**Certificate course in Big Data, Grade: A, 2014**

**UNIVERSITY AT BUFFALO, THE STATE UNIVERSITY OF NEW YORK, Buffalo NY**

2007-2008

**Master of Science in Management Information Systems, GPA: 3.9/4.0, June, 2008**

**NATIONAL INSTITUTE OF TECHNOLOGY (Formerly REC), Durgapur, India**

2001-2005

**Bachelor of Engineering in Computer Science & Engineering, GPA: 3.8/4.0, May, 2005**

*Graduated with Honors: Rank# 3 out of 66*

*NIT's Merit Scholarship Holder, 2003-2004*

*Everready Scholarship for all 4 years, 2001-2005*

---

## CORE COMPETENCIES

---

- Generative AI and LLM Architectures (Fine-tuning, Distillation, Function Calling)
- Multi-Agent Conversational AI Frameworks
- Scalable ML Infrastructure (Kafka, Spark, Kubernetes, Docker)
- Engineering Leadership for Developer-Focused AI Platforms
- Cross-Functional Collaboration and Technical Evangelism
- Patented Innovation in NLP, ASR, and Multi-Modal AI
- Real-Time Data Processing and Edge AI Systems

### References:

Available upon request.