

# SHARAD VARSHNEY

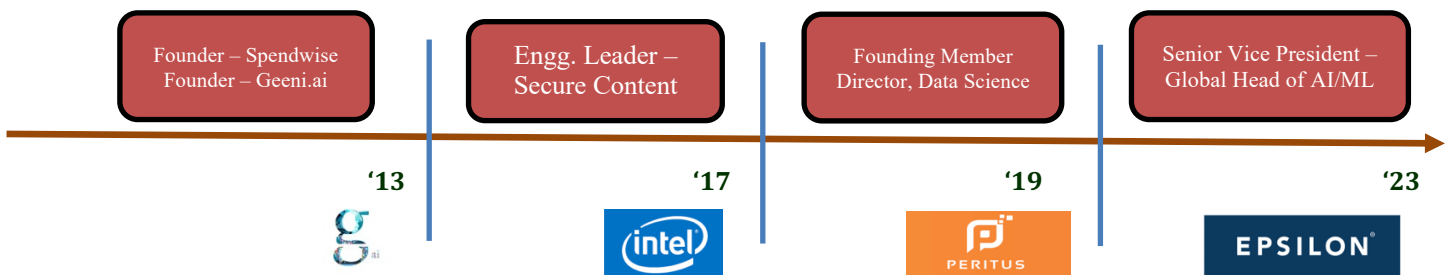
650-666-7984

Research and results-driven, hands-on influential, Deep Learning, Generative AI leader with more than 15+ years' experience and unique cross domain advanced level experience in Generative Artificial Intelligence, advanced machine learning and Deep Learning with training and fine-tuning large language models world of products, who drives growth by driving insights that fuel increased profitability and strategic transformations. Strategy and insights driven professional who builds and leads productive teams, drawing on years of progressive, cross-industry success on an international scale. A decisive and resourceful data science and technology leader who leads multiple ML/AI/Data/Engg teams and skilled at cultivating business alliances and establishing development and operational software engineering policies.

## AREAS OF EXCELLENCE

- ◆ Generative Pretrained Transformers ◆ Generative AI ◆ Artificial Intelligence ◆ LLMs ◆ Deep Learning
- ◆ Model Serving / Deployments & Migrations ◆ Docker & Kubernetes ◆ Next Best Action
- ◆ Cloud & Data Center Operations ◆ Customer Intelligence Data Platform
- ◆ Training and Scoring at scale ◆ A/B Model Testing

## CAREER JOURNEY



## TECHNICAL SKILLS

|                 |   |
|-----------------|---|
| LLM             | GPT-4, LLaMa2, MPT, Falcon, Dolly-v2-7b, T5, GPT-2, Knowledge Graph               |
| Big Data / ML   | Spark ML, Tensorflow, PyTorch, Keras, PySpark, SpaCy, Weka, Apache Hadoop – Spark |
| Prog. Languages | Python, C#, Java, C, JS, Sql, Pl/Sql  |
| Web framework   | MEAN, MERN, Flask, Spring Boot, Python with Flask, ASP MVC / Web API.             |
| ML Algorithms   | Logistic, Linear, SVM, GBMs, DNNs   |
| Databases       | AWS RedShift, PostgreSQL, Mongo, Oracle, SQL Server                               |

## RESEARCH WORK

**Micro Moments Identification:** Micro-Moment analysis and identification are set of Bayesian network algorithms that determine accurately in real-time across the omni-channel customer journey the exact moments, channel and content to engage people with cognitive experiences curated programmatically in a prescriptive engagement.

**CORD19- COVID-19** – Designed and developed open-source search engine indexing for more than 50M research articles- [deliver Question/Answer model based on BERT and Solr](#).

**Calamity (COVID-19) Index feature:** This research creates an index feature to accurately help predict also bridges the gap for demand forecast during COVID or any other socio-economic disasters and used to predict different product demands in the region.

**CustomerBehavior2Vec:** Vector embeddings created for more than 250 million consumers based on their demo and socio graphic attributes

## EDUCATIONAL QUALIFICATIONS

### Deep Learning Specialization,

Neural Network and Deep Learning  
Structuring Machine Learning Projects

### Applied Machine Learning,

By Andrew Ng

**M.S. in Computer Science,**

**B.S. in Computer Science**

### Coursera

Improving Deep Neural Networks: Hyper parameter tuning, Regularization Optimization  
Convolution Neural Network/Neural Networks and Deep Learning

Deep Learning – NN/Sequence Models

### Stanford Online

**CSUS, CA**

**NMU, India**

## PROFESSIONAL EXPERIENCE

**Principal Data Scientist**  
**Sr. Vice President, Epsilon**  
**Remote - San Francisco, CA**

**Duration: May' 19 – Current**

**Description:** Epsilon is built on a succession of innovations and transformations that have solidified the culture and expertise that continue to run deep throughout our parent Groupe. Epsilon one of the Groupe companies, is the leader in outcome-based marketing. Epsilon has a rich, 50-year heritage in helping marketers anticipate, activate, and prove measurable business outcomes. We create data-driven, enriched customer experiences focused on real people, with 200M + unique individuals, sending 100B + emails, \$15 trillion transactions tied to people, Epsilon People Cloud Customer, Loyalty, Prospect platforms have unique offerings.

### **Responsibilities:**

- Collaborated with cross domain leadership teams such as R&D, Product Mgmt., Engineering teams to drive product innovation and achieving a 45% increase in overall productivity through strategic and thought leadership, effective model development lifecycle management.
- Managed and led R&D team on fine tuning llama2, Blender Bot, and open-source LLMs, and used them for various purposes such as Data Enrichment, Data Cleansing, Data Analysis, Campaign Segments filtration and creation purposes in clean room solutions such as Epsilon People Cloud Prospect.
- Led a team to develop LLMU, BBH, AGI eval, Math and RC datasets evaluated on LLMs.
- Currently leading a team to fine-tune llama2-7b model on proprietary datasets, so it can answer logical reasoning and product related questions with good confidence.
- Innovated and adding new directions to team for R&D work in Speech-2-Text, based on **CitriNet**, and Conformer CTC and Whisper AI based deep neural network transformers models to convert audio data into text.
- Fully owned, designed, and developed and led geographically distributed global teams responsible for developing data distributed based Campaign Orchestration Decision Engine with machine learning models:
  - **Next Best Action** based on Reinforcement Learning, identifies next best action for the customer
  - **At Risk** – Predicting future attrition based on innovative approach using Sequence Embeddings – GRU to predict 4-12 months out in future.
  - **Lifetime value** – Predicted Customer Lifetime value predictions in future 1,3 and 6 months.
  - **Recommendation Engine** to identify next 5 SKU level products customers might be interested, post purchase.
  - Text entity extractions from images to extract texts from natural scene, using OpenCV, and Tesseract libraries.
  - Designed Customer lifetime journey (CLJ) capability to stitch events as micro moments in CLJ.
  - Enable 1-on-1 personalization for retail customers and enable monetization of data assets including Total Source Plus MarketView (190M+ US Household demographics and market spend behavior data)
- Innovated, researched and Productionalized COVID-19 index and calamity index features to drive product propensities models and demand forecasts models and product demands per channel.
- Responsible for ML capability delivery for PeopleCloud Customer (a CDP implementation) and ML and data driven intelligent customer segmentations using Customer Lifetime Value, Customer Churn, Propensity models.
- Inspired, led and grew a team of 50+ data science personnel including Directors, Managers, Data Scientists and Data Engineers.
- Brought together a team of versatile and top-notched Data Science team, hiring from renowned institutions such as MIT/Columbia/CMU and world-renowned companies like Google, Apple and Microsoft.
- Managed a broad and diverse set of stakeholders, setting expectations, and working collaboratively at multiple levels of the organization to get things done.

**Environment: Generative AI, Transformers, Python, PyTorch, Databricks, Spark, Deep Learning.**

**Founding Data Scientist, Peritus.ai,**  
**Palo Alto, CA**

**Duration: Oct'17 – May' 19**

**Description:** Peritus.ai has a vision for Autonomous Data Centers and PeritusAgent is virtual expert system for assisting and automating support fulfillment for engineered IT systems in enterprise datacenter and is designed for product vendors and affiliated service partners. It offers 1. Curation and consolidation of product knowledge , vendor validated design and solution deployment practices, 2. Automation of decision making across the service management chain through cognitive modeling of incident resolution and service fulfillments based on: modeling past incident resolutions, incident related feature extraction from system logs, 3. Session-based training from SMEs, 3. Abilities to deliver new system entitlements through full automated system fulfillment in order to deliver fully automated incident classification and routing and normalized logging, smart orchestration of incident diagnosis, resolution and recovery between SMEs, quantifiable business impact in fulfillment of service entitlements.

### **Responsibilities:**

- Formed vision for startup to solve incident based tickets in automated fashion using set of simple heuristics and ml models

inference in integrated fashion to auto-route tickets to appropriate teams, apply application KB articles to identify optimum configurations and be able to apply those configs to h/w.

- Managed and Led team of engineers responsible for running **200 node Spark cluster** with more than 20 TB memory, for cleaning up huge amount of time series data.
- Designed, and developed successfully deliver set of language models using Tensorflow based ML pipeline
- Designed and developed Language models:
  - Developed language-based entity parsing, entity extraction models for understanding texts and finding similarities.
  - Deep NN using multi-layer Stacked LSTMs and Conv layers to reduce model hyperparams and extract features.
  - NER model - Character embedding using Bidirectional LSTM with CRF
  - Sup. learning models for problem categorizations for smart routing
  - Sequence to Sequence – Generative models
- Developed model for detecting anomalies on millions of logs entries by indexing timestamps and timeseries construction.
- Developed custom model serving framework with similarity computation for text-based incidents data.
- Designed complete Information Retrieval pipeline to ingest data and implemented custom idf algorithm in Solr to make search results more relevant.
- Converted product vision and product capabilities in set of technical architecture and deliverables for the team.
- Currently working on creating Neural Network based Recommender System – which will create recommendation for PeritusAgent and become the core product of Data and ML ingestion pipeline.
- Worked on Storage fillers, Cisco switches, load balancers-based hardware devices for Cisco, Citrix, HPE, Aruba to engineer ML features and apply ML-AI deep neural network models to these use-cases.
- Fully responsible for all search related Indexing, Crawling, ranking optimization in Apache Solr.

**Environment: Python, Tensorflow, Java, Spark, Kafka, Athena, Redis, Keras, Apache Solr, AWS, HDFS.**

**ML Engg. Lead,  
Intel Corp,**

**Duration: Jun'12 – Oct'17**

**Intel Corporation** is a multinational semiconductor chip maker corporation headquartered in Santa Clara, California. Intel is the world's largest and highest valued semiconductor chip maker. It is the inventor of the x86 series of microprocessors, the processors found in most personal computers. Intel Corporation, founded on July 18, 1968, is a portmanteau of Integrated Electronics (the fact that "intel" is the term for intelligence information was also quite suitable). Intel also makes motherboard chipsets, network interface controllers and integrated circuits, flash memory, graphic chips, embedded processors and other devices related to communications and computing.

**Project # 1 – Secure Content Pipeline** with Resource Design Center with top picks based on recommendation engine.

**Project # 2 – ML - Anomaly Detection – Canary Deployments**

**Project # 3 – ML Based – Incident classifications** and led multiple teams and projects for successful delivery.

**Founder,  
Spendwise, Inc.**

**Duration: Jun' 11 – Nov'12**

**Founded Fin-tech startup based on consumer spending patterns and successfully implemented product.**

**Lead Engineer  
GTS, San Francisco, CA**

**Duration: Sep'06 – May'12**

**Managed and lead multiple software engineering teams on financial and healthcare product apps, worked on multiple projects from full stack web projects to Big Data and distributed pipelines.**

*More details on projects and references can be made available on request*