

## Juby Jose/ Director – AI at Cactus Labs”

- 19+ yrs. track record of hands-on Technology leadership, working at the intersection of applied AI research, real customer applications and cloud products to enable it
- Built AI based digital editing service, capable of thoroughly editing entire research manuscript
- Built AI engine which understands users reading interests deeply and finds relevant topics that otherwise are easy to miss out.
- Built Apollo machine learning platform (now Wipro HOLMES for Business).
- Built Semantic Data Hub platform – Build and operate enterprise Brain.
- Built AI Hub – One stop hub for teams involved in training, tracking and deploying machine-learning models.

 Jubyjose2007@gmail.com

 +91 - 8105132132

 Bangalore/Mumbai

### Core Competencies

Machine Learning Engineer
Product Development
Customer Acquisition / Pre-sales
Hands-on Innovator
Trusted advisor for strategic customers
Product Management
Technology consulting
Digital Strategy
Team Building & Leadership

### International Work experience

- Nortel Networks, Ottawa, Canada(1.5yrs)
- SFR, Paris, France (11 Months)
- Holds US B1-visa

### Trainings

- Data engineering, Big Data and Machine learning on **GCP, AWS and Azure**
- Natural Language Processing Nanodegree
- Data Structures & Algorithms Nanodegree
- Machine learning , Deep learning by Andrew Ng(Coursera/Deeplearning.ai)
- Hands-on training in AI and accelerated computing from NVIDIA DLI.

### Professional Experience

**Cactus Labs**, Mumbai, India Jan 2020 - Present,  
*Director – AI/ML, 3.4 years*

**Wipro Ltd**, Bangalore, India Sept 2004 - Dec 2019,  
*Engineering Director & DMTS, 15.3 years*

### Academic Details

- MS Software Engineering - BITS Pilani
- (Computer Applications, Mathematics, Statistics) - Kuriakose Elias College, Kerala

### Patent Details

- AI based supplier recommendation – *Juby Jose, Roman Kuzmik*

### Research

- Structured Pruning Adapters

## Skill Set

### DEEP LEARNING SKILLS

- Natural language processing
  - Transformers, BERT, RoBERTa, T5, GPT2,GPT3.5, GPT4
  - NLTK, spacy, Stanford CoreNLP, Stanford Stanza
  - Azure LUIS, BOT builder, Rasa
- Deep Learning for Computer Vision
  - Video understanding and prediction
  - Human activity recognition
  - Automated aircraft inspection
  - R-CNN, YOLO, Mask R-CNN ...

### Programming Languages

C++ | Java | R | Python

### Deep learning Frameworks

Tensorflow | Keras | Hugging Face| PyTorch |  
Numenta-HTM

### Datascience ISV's

H20.ai | Iquazio | DataRobot

### Public cloud expertise

AWS | Azure | GCP

### BigData platform expertise

Cloudera | Databricks| Hadoop | Spark

### Open Source software products

Apache Eagle | Apache Griffin | Apache  
predictionio | Apache Airflow| Apache Beam |  
Kubernetes | Horovod | clipper.ai | Kubeflow |  
MLflow | Weights & Biases | TensorFlow Serving  
| TensorRT | NVIDIA/DIGITS

## ACCOMPLISHMENTS

19 years	Competence			Role	Key contribution
3.4 yr. 2020	Cutting edge Technology Skills	Innovation Skills	Artificial Intelligence	Director AI/ML, <a href="#">Cactus Labs</a>	<ul style="list-style-type: none"> <li>Co-created and built large deep learning models for <a href="#">grammatical error correction(GEC) systems</a>, paraphrasing, text-infilling, autocomplete etc.</li> <li>Built academic language models (BERT, RoBERTa, T5, Vicuna/LLaMA) using scientific text.</li> <li>Co-designer and implementer of sever internal tools/frameworks for large scale data preprocess and end to end machine learning life cycle management (MLOps).</li> <li>Creating AI roadmap for the Cactus Labs</li> <li>Leading AI team at Cactus Labs – pursuing applied research in AI and making innovative products for the advancement of science.</li> </ul>
1 yr. 2019				Distinguished Member of Technical Staff(DMTS) & Engineering Director at <a href="#">Wipro</a>	<ul style="list-style-type: none"> <li>Business growth strategies through new products, new services and innovation through AI technologies</li> <li>Creating technical roadmap for the SL</li> <li>Leveraging expertise to create opportunities and close deals</li> </ul>
1 yr 2017		Customer Consulting Skills		AI practice Leader, DPE.ai	<ul style="list-style-type: none"> <li>Practice AI business roadmap and strategy</li> <li>Achieved continuous business growth.</li> <li>Creation of innovative, high quality solutions</li> <li>AI solutioning, pre-sales.</li> <li>Managed pre-sales functions, developed proposals, prepared &amp; delivered presentations to prospective customers and achieved revenue targets</li> </ul>
1yr 2016				AI Consulting, DPE.ai	<ul style="list-style-type: none"> <li>Strategic Consulting AI and ML</li> <li>Incubating AI COE's, Co-creation, and end-to-end machine learning life cycle management.</li> </ul>

1 yr 2015				<b>Engineering Manager, Apollo platforms</b>	<ul style="list-style-type: none"> <li>Product Management- from inception through GTM and execution</li> </ul>
2 yr 2013				<b>Product Architect &amp; AI researcher, Apollo Platforms</b>	<ul style="list-style-type: none"> <li>Solution architecture, development and delivery.</li> <li>Applied AI research in partnership with IIIT-B.</li> </ul>
2 yr 2011				<b>Architect /Engineering Manager</b>	<ul style="list-style-type: none"> <li>Solution architecture and implementation.</li> <li>Increase 2M revenue through CRs.</li> <li>Worked from client location in France(Paris)</li> </ul>
2 yr 2009				<b>Technical Lead</b>	<ul style="list-style-type: none"> <li>Group Head for Billing manager application</li> <li>Worked from client location in Canada(Ottawa)</li> </ul>
2 yr 2007				<b>Senior Software Engineer</b>	<ul style="list-style-type: none"> <li>Feature enhancements for billing manager Application.</li> <li>Worked from client location in Canada(Ottawa)</li> </ul>
3 yr 2004				<b>Software Engineer</b>	<ul style="list-style-type: none"> <li>Application support and maintenance</li> </ul>

## Work Experience

Led key strategic initiative with high impacts across the organization. Involved in driving product roadmap strategy, production architecture, product development, customer acquisition and engagement. Managed stakeholder across levels within prospects & customers around value articulation (engaged from prospecting to deployment).

1. [Paperpal](#) – Fast and flexible editing tool used for all academic writing
  - Paperpal provides instant suggestions for language improvements while writing a manuscript.
  - Built large deep learning models trained on billions of editorial corrections (both the original and edited files) performed by our industry-leading board of editors, to provide suggestions for improvement on par with those provided by a human editor.
  - Corrected grammatical errors and improve phrasing and word choice to improve the quality of the language
2. [R Discovery](#) – *Research discovery and* notifications/email alerts to ensure researchers always have the recently published research content at their fingertips.
  - R Discovery is a free app for students, researchers and others in academia to find and read research papers.
  - Built models to extract concepts automatically from scientific books chapters or journals.
  - Built personalized recommendation models using user interaction dataset, articles metadata, journal ranking/metrics.

- Built larger scale semantic search engine with millions of research articles using AWS Opensearch and sentence transformers.
3. **AI Hub - Build and operate end-to-end machine learning life cycle:** *One-stop hub for teams interested in training, tracking and deploying AI models at scale. Feature includes:*
    - *Version control data similar to what git does for code*
    - *Provide collaborative environment for experimentation service*
    - *Infrastructure agnostic - any cloud, hybrid, on-premise*
    - *Flexibility to use ML tools/frameworks of your choice.*
    - *Use Docker containers with Kubernetes and other existing tools to scale model deployment*
  4. **SDH – Semantic Data Hub:** *Enables enterprises to ingest both structured as well as unstructured data sources, extracts semantic meaning and relationships, and represents knowledge in a computable form that would enable whole new classes of intelligent applications powered by AI and curated knowledge. Examples of these applications include*
    - *Supplier capability analytics*
    - *Real world evidence for pharma.*
    - *Question-answering services*
    - *AI assisted product analyst.*
    - *Knowledge-powered bots*
  5. **Apollo Anomaly Detection Platform:** *Apollo leverages Machine Learning & Big Data analytics to identify various frauds & compliance violations across industry segments. The product is currently in production in multiple process areas including payments, insurance claims & other areas of compliance. (Combination of unsupervised & supervised learning ensemble algorithms).*

### **Key solution developed for various customers across industries:**

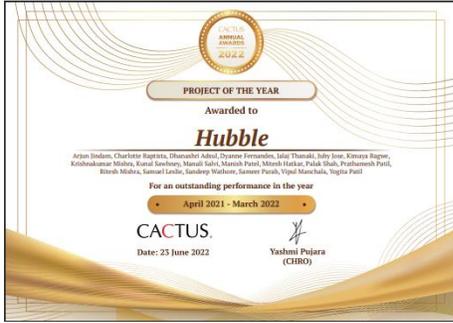
- a. **Supplier Capability Assessment and Management:** *Leveraged Enterprise Knowledge Graph capabilities allowing customer to enhance acquisition, retention and application of the supplier assessment information across organization. This includes:*
  - *Illustrating dynamic relationship between siloed information across the organization.*
  - *Better understanding of its supplier capabilities.*
  - *Organize and distribute the organization's knowledge accessible to every authorized employee.*
  - *Make decision based on comprehensive knowledge.*

- b. **Intelligent Edge:** *Build and operate an intelligent edge-computing platform for real time video surveillance, human pose/activity detection. This includes:*
- *Build and operate a low latency prediction serving system at edge..*
  - *Alert on suspicious activity or face similarity matching detection within 1 second*
  - *Process multiple video streams with throughput of 30 frames per second for each video stream*
  - *Handle Model Versioning , Model Life Cycle Management and perform A/B testing & rolling upgrades of the Models & ML runtimes*
  - *Works with any ML library, language & existing code.*
  - *Deploy the ML models on cloud as well as at the edge.*
- c. **AI assisted Product Analyst:** *Build and operate an intelligent semantic hub connecting billions of data records pertaining to millions of mobile devices for data exploration, failure analysis, pattern detection, trends and recommendation services. This includes:*
- *Ability to automatically identify the furthest most step in the supply chain that likely has the highest defect density around it, potentially identifying high risk devices before they are sold*
  - *A visual interface for analytics of inflow and outflow from that step in the supply chain along all of its dimensions, such as vendor, week of manufacture, site, etc.*
  - *Custom graph based anomaly/pattern detection on top of the Supply Chain knowledge graph.*
  - *Accelerate the discovery of systematic problems in millions of phone activated across geographies with various components manufactured by different vendors having multiple plants.*
  - *Search for relevant areas/modules of high defect density in the Supply Chain of the phone*
  - *Compare selected supplier vendors manufacturing similar modules across Non-causal dimensions like region, carrier etc.*
  - *Recommend root cause of the supply chain problem using AI/ML on top of Supply Chain knowledge graph with causal inference*
- d. **AIOps - IT Operational Analytics:** *Built an enterprise grade proactive, self-healing system using AI and NLP technologies, to reduce downtime and increase customer satisfaction. This includes:*
- *Created an aggregated & correlated view from multiple disparate monitoring systems and predict/remediate failures in advance.*
  - *Introduce real time Predictive Analytics for middleware infrastructure monitoring based on events, metrics, logs, etc.*
  - *Perform remediation based on various predictions*

---

## Recent recognitions

## PROJECT OF THE YEAR, Cactus



## DMTS felicitation by K R Sanjiv, CTO

