

# Arun Jayapal, Ph.D.

AI Architect — Computational Linguist — LLM Specialist

📍 Tamil Nadu, India    ✉ jayapala@tcd.ie    🌐 github.com/arunjeyapal    📧 arunjeyapal.github.io

**Research-backed Data Scientist and NLP Architect** with over a decade of experience bridging the gap between theoretical Bayesian modeling and production-grade Generative AI. Expert in designing agentic workflows, deploying LLMs (GPT-4, Llama 2) for specialized domains, and optimizing NLP micro-services. Proven leadership in AI Quality Assurance and algorithmic innovation across LegalTech, Fintech, and Speech Analytics sectors.

## EDUCATION & RESEARCH

### Trinity College Dublin (University of Dublin)

Dublin, Ireland

*Ph.D. in Computer Science (Computational Linguistics)*

2013 – 2018

- **Thesis:** "Finding Diachronic Sense Changes by Unsupervised Methods." Proposed a probabilistic generative model to identify emerging word senses over time using Expectation Maximization and Gibbs sampling.
- **Grant:** Secured €88,000 research grant from Science Foundation Ireland (CNGL/ADAPT).
- **Teaching:** Tutored "Advanced Computational Linguistics" (Unsupervised Machine Translation & Topic Modelling).

### The University of Sheffield

Sheffield, UK

*M.Sc. in Computer Science (Speech & Language Processing) – First Class*

2010 – 2011

- **Focus:** Named Entity Recognition (HMM/CRF) and Speech Intelligibility enhancement.

## TECHNICAL ARSENAL

- **Generative AI & LLMs:** GPT-4, Llama 2/3, LangChain, Flowise, Pinecone (Vector DB), Prompt Engineering, RAG Architectures.
- **Machine Learning & NLP:** Bayesian Modelling, CRF/HMM, Scikit-learn, PyTorch, NLTK, Stanford CoreNLP, Speech Analytics (Thrax).
- **Data Engineering:** PostgreSQL, Clickhouse, Elasticsearch, Airflow, Celery.
- **DevOps & Backend:** Python (FastAPI/Flask), C++, Docker, AWS (EC2 cost optimization).

## PROFESSIONAL EXPERIENCE

### Freelance Consultant

Remote

*AI Architect & NLP Specialist*

Jan 2024 – Present

- **Multi-Modal Video Audit Platform:** Developed a video analysis engine for influencer marketing that leverages Multi-Modal LLMs to conduct automated audits of ad content, assessing visual and audio streams for brand compliance and performance.
- **Synthetic Agentic Systems:** Engineered a simulation environment using synthetic agents to validate product ideas and market strategies. The system utilizes autonomous multi-agent interactions to generate rapid feedback and scenario testing.
- **Generative AI Solutions:** Architected a summarization bot to filter emails and draft responses automatically using Agentic workflows.
- **Domain-Specific QA:** Developed a pharmaceutical research chatbot utilizing GPT-4 and Pinecone DB for semantic search.
- **Social Analytics:** Built sentiment analysis pipelines for Instagram influencer marketing using PaLM (TextBison) and GPT-3.5.

### Onit India Pvt Ltd

Pune, India (Remote)

*Data Scientist*

Sep 2022 – Dec 2023

- **LegalTech LLMs:** Experimented with Open-Source LLMs (Llama 2) for Legal NER and issue identification in contracts.
- **Optimization:** Optimized LLM deployment strategies, reducing EC2 instance costs while maintaining response latency.
- **Predictive Modeling:** Developed regression models for cost estimation on block-billed legal items.

### EY (Ernst & Young) GDS

Bangalore, India

*Assistant Director*

Nov 2019 – Aug 2022

- **AI Governance:** Established an AI Quality Assurance framework focusing on Robustness, Fairness, and Explainability.

- **Document Intelligence:** Built an end-to-end system for assigning Harmonized Trade System (HTS) codes using multi-label deep learning classifiers.
- **Innovation:** Awarded "Pursuing Innovation Award" (2020) and "Exceptional Client Service Award" (2022) for breakthrough implementations in document intelligence.

## Sayint.ai (Zen3 Info Solutions)

Senior Software Engineer

Hyderabad, India

May 2017 – Oct 2019

- **Speech Analytics Platform:** Engineered a comprehensive analytics engine detecting emotion, monologues, and speaking rates from dual-channel audio.
- **PII Redaction:** Integrated NER engines with Celery to identify and mask personal information (PII) in real-time calls for compliance.
- **Expert Systems:** Designed rule-based and supervised classifiers to score agent script adherence and sentiment in the travel domain.

## EARLY CAREER HIGHLIGHTS

---

- **MagnaInfotech (Emirates Client):** Developed outlier detection models to predict aircraft spare part failure and "Aircraft on Ground" (AoG) situations.
- **IB Technology (Nuance):** Led the transition from rule-based to statistical NLP for Voice-to-Text services (English, Spanish, French).
- **Evalueserve:** Conducted deep-dive patent analytics and invalidation searches for wireless and encoding technologies.

## THOUGHT LEADERSHIP & INVITED TALKS

---

- **IIT Jodhpur:** Delivered an expert session on **AI Governance**, covering regulatory frameworks, model fairness, and the ethical deployment of Generative AI.
- **Industry Speaker:** Frequent speaker at various technical forums on the topics of Large Language Models, NLP Architectures, and AI Safety.

## SELECTED HONORS

---

- **Innovation Award (EY, 2020):** For implementing novel HTS code classification.
- **Research Grant (SFI, 2013):** Full scholarship and stipend (€88k total) for Ph.D. research at Trinity College Dublin.