

ARPAN GHOSHAL

✉ contact@arpanghoshal.com 📞 +91 96678-05577 arpanghoshal.com linkedin.com/in/arpanghoshal

Work authorization: UK HPI visa (no sponsorship) | Netherlands Zoekjaar permit (no sponsorship)
US approved H-1B (transfer only) | **Available Immediately**

PROFILE

Applied AI and Forward Deployed Engineer with **5+ years** of experience building production AI systems: AI agents, LLM applications, RAG, evals and guardrails, NLP, and machine learning. Works directly with founders and customers, taking ambiguous problems from scoping and architecture through implementation, evaluation, deployment, and handover. Most recently upgraded production AI systems for **3 startup clients** across Europe and the US, and before that was Director of AI Innovation for Halsey Minor's ventures, leading a team of 12 and building two 0-to-1 AI products end to end. Creator of **EmoRoBERTa**, an open-source NLP model with **15M+ downloads and 100+ citations**, and of **ctrlrun**, an open-source execution-safety layer for AI-agent actions and an **official LangChain middleware integration**. Has worked with founders across fintech, HR tech, utilities, neuroscience, legal and news.

EDUCATION

New York University (NYU), New York, USA 2022 – 2023
MS in Computing, Entrepreneurship and Innovation, Courant Institute of Mathematical Sciences & Stern School of Business
Completed the two-year program in one year; research assistant in deep learning and applied AI.

PES University, Bangalore, India 2017 – 2021
BTech in Computer Science and Engineering, Data Science Specialization

EXPERIENCE

Applied AI Engineer (Independent Consultant) Mar 2026 – Sep 2026
Independent AI Engineering & Consulting, Europe & United States, Remote

- Engaged by **3 startup clients** across Europe and the US to upgrade production AI systems, working directly with founders from scoping through architecture, delivery and handover.
- Audited and re-architected existing LLM and RAG pipelines, **reducing inference and infrastructure costs** while **improving answer accuracy and system efficiency**.
- Built **evaluation suites** (golden datasets, LLM-as-judge scoring, regression tests) to benchmark each client's AI features, then used the results to ship substantial feature improvements.
- Delivered AI agents, MCP servers and workflow automation on top of the upgraded stacks.
- Retained by Operating Equity Partners** at the founder's request to maintain and extend the AI systems built for their ventures.

Director of AI Innovation Feb 2025 – Feb 2026 · retained to Sep 2026
Operating Equity Partners (Vivid Labs Holdings, Inc.), San Carlos, CA, Remote

- Led an **international team of 12 across design, product and engineering**, running the company end to end and partnering directly with **Halsey Minor (co-founder of CNET, Salesforce, Google Voice and NBCi)**.
- Built **both 0-to-1 AI products end to end**, a trusted news network that fights misinformation and an AI avatar legal assistant, owning the architecture, LLM pipelines, retrieval, backend and integrations.
- Secured **continued investor funding** for the ventures by demoing the working products built.
- Built the **agent reasoning layer** for legal and news-verification workflows: multi-step investigation, evidence retrieval, tool-augmented reasoning and typed structured outputs.
- Designed the **eval and observability stack**: golden test sets, LLM-as-judge and human-review scoring for factuality and hallucination, output guardrails, and LangSmith/Langfuse tracing of quality, latency and cost per request, used to catch regressions before every release.
- Fine-tuned open-source LLMs for legal and news use cases and shipped them as product features.

Machine Learning Engineer Sep 2023 – Feb 2025
Gisual Inc., Philadelphia, PA

- Built the **first commercial machine learning model for power outage detection**, achieving **96% accuracy** and owning the pipeline from geospatial data acquisition and feature engineering through deployment for utility clients.
- Designed a custom outage-detection architecture combining mathematical modeling, deep learning, and ensemble methods to capture relationships across geospatial, weather, grid, infrastructure, and device data sources.
- Built forecasting models to predict outage patterns and demand trends, improving proactive resource allocation for utility clients.
- Designed and shipped a **GenAI-powered outage detection system** that analyzed social media signals in real time to identify outage events ahead of conventional monitoring systems.
- Deployed solutions across **US and international markets** with monitoring dashboards, automated retraining workflows, and foundations for a fiber outage detection model.

- Architected an AWS-based machine learning pipeline using SageMaker, Lambda, and S3 for 24/7 web crawling, ingestion, and large-scale text preprocessing.
- Developed a BERT-based semantic matching model for skills-to-context mapping, improving matching accuracy by **170% over baseline** for the workforce intelligence product used by enterprise talent teams.

- Built and deployed a real-time AI-powered audio analysis API serving **150k+ users**, training NLP models to extract 20+ engagement metrics.
- Applied Canonical Correlation Analysis to EEG and audio stimulus data for the company's neuroscience product.

- Built the company's **first data infrastructure and data APIs** from scratch; later adopted by **major fintech clients** on the platform.

TECHNICAL SKILLS

Agents & LLMs	Claude Code (power user), MCP servers and gateways, LangChain, LangGraph, OpenAI Agents SDK, OpenClaw, OpenAI/Anthropic APIs, tool calling, structured outputs (Pydantic), RAG, Pinecone, ChromaDB
Evals & Safety	Evals, guardrails, human approval gates, agent audit trails, LangSmith, Langfuse, OpenTelemetry
ML	PyTorch, Hugging Face Transformers, fine-tuning (LoRA/QLoRA), scikit-learn, XGBoost, TensorFlow
Languages & Backend	Python, TypeScript, FastAPI, Flask, REST APIs, WebSockets
Cloud & Data	AWS (SageMaker, Lambda, S3, EC2, ECS/EKS, RDS, Step Functions), Cloudflare, Docker, Kubernetes, Terraform, GitHub Actions, PostgreSQL, Redis, MongoDB, Elasticsearch, MLflow

PROJECTS

EmoRoBERTa (Open Source, Creator) · 2020 RoBERTa-based emotion detection model; top trending emotion model on Hugging Face (2020–2022) with **15M+ downloads and 100+ citations**; published at IEEE CONECCT 2022. huggingface.co/arpanghoshal/EmoRoBERTa

ctrlrun (Open Source, Creator & Maintainer) · 2026 Execution-safety layer for AI-agent actions: binds a human approval to the exact action, runs each effect at most once across hosts, treats an unknown outcome as ambiguous rather than failed, and leaves a hash-chained receipt. **Official LangChain middleware integration**; MCP gateway, LangGraph and OpenAI Agents adapters. Python, SQLite/Postgres. v0.12 on PyPI. github.com/CTRLRun/ctrlrun | ctrlrun.dev

ctrl ai agents · 2026 Hosted control room for every AI agent a company runs, bought or built (Claude Code, Cursor, ChatGPT, LangChain, anything on MCP): records every action, applies allow, ask or block policy before it runs, and flags drift. ctrlaiagents.com

ctrl payments · 2026 The same boundary for money: refunds, vendor payments and payouts an agent attempts go through, wait for a person to approve that exact amount, or are refused, with per-agent budgets and split-payment detection. ctrlpayments.com

ContractReviewer · 2025 LLM tool that flags contract risks, hidden fees, and one-sided clauses in plain English; **PandaDoc's CEO and CTO** proposed integrating it into their platform. contractreviewer.io

iwrote · 2026 Certifies writing as human. Records how the words arrive in the browser (typing, edits, pauses, pastes) and gives the writer a public certificate link instead of an AI detector's guess. iwrote.co

Puchku · 2026 Privacy-first wellbeing companion for Mac, iPhone and iPad. Runs on-device AI, including speech-to-text; end-to-end encrypted sync with device-only keys, no accounts. Launching soon. puchku.app

Video Frame Prediction (NYU Deep Learning by Yann LeCun) · 2022 Semi-supervised model predicting frame 27 of a clip from its first 13, via contrastive learning on unlabeled video.

PUBLICATIONS & STANDARDS

- [1] Section author, "Non-repudiation, Immutability, and Retention," **OWASP GenAI Security Project, Agentic Security Initiative**, 2026. Five-level evidence model for agent audit trails, from append-only to signed.
- [2] "An Enhanced Context-Based Emotion Detection Model Using RoBERTa," in *Proc. 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, 2022, pp. 1–6.
EmoRoBERTa: top trending emotion model on Hugging Face (2020–2022). 100+ citations, 15M+ downloads. huggingface.co/arpanghoshal/EmoRoBERTa
- [3] "CATAQ: Concise Answer to Any Question," in *Proc. 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICES)*, 2021, pp. 1–7.

CERTIFICATIONS & RECOGNITION

AWS Certified Solutions Architect | NASA Space Apps Challenge Mentor & Judge