ARPAN MISHRA

Applied AI Engineer · Gen AI & Agentic Systems
New Delhi · +91-9958325792 · mishraarpan6@gmail.com · Portfolio

SUMMARY

Applied AI Engineer with ~5 years of experience building production-grade Gen AI solutions, agentic systems, and copilots for enterprise clients in regulated environments. My technical expertise spans multi-agent orchestration, retrieval-augmented generation, intelligent document digitization, LLM evaluation and prompt management at scale, and full-stack AI integration — across domains including clinical trial documentation, healthcare AI, document intelligence, and regulated AI deployment. I approach every problem by questioning the problem statement first, identifying the real bottleneck before committing to a solution. I am equally comfortable going deep on system architecture and implementation details with engineers, and communicating the same decisions as business impact to real users & business stakeholders. I am most energized at the 0-to-1 stage: taking ambiguous, complex problems and driving them all the way to production.

CORE SKILLS

Gen AI: Multi-Agent Orchestration, RAG, Document Digitization and Parsing, MCP Server Design, LangGraph, LangChain, Claude SDK, LangSmith, Portkey, CoT / Few-Shot / Structured Prompting, LLM Evaluation (LLM-as-a-Judge, ACS Framework), State & Memory Management
Cloud & Infra: AWS (Bedrock, Bedrock Data Automation, SageMaker, Agent Core, EKS, DocumentDB), Google Cloud (Vertex AI, Gemini API), Azure Document Intelligence, Docker
ML: NLP, Layout Detection, spaCy NER, HuggingFace, PyTorch, XGBoost
Languages: Python (expert), SQL, R, Bash

WORK EXPERIENCE

Data Science Consultant | ZS Associates
Dec 2025 – Present · New Delhi

Clinical Trial Document Authoring Platform (Fortune 500 Pharma · MS Word Add-in Copilot)

- **Problem:** Medical writers spent hours synthesising 100+ regulatory documents to produce a first draft — documents reviewed by the FDA where even minor inaccuracies trigger rejection and costly rework. Worked with **Medical Writers, Clinical Scientists, PMs, and Solution Architects** to build a copilot embedded in MS Word that reduces first-draft time to minutes.
- The core challenge was standardising information across 100+ varied regulatory formats into a structure an LLM could reason over accurately. Built a **multi-mode digitization pipeline**: fine-tuned **Detectron2** (pretrained on PubMed) for client-specific document layout detection; enhanced **AWS Bedrock Data Automation + Azure Document Intelligence** with custom post-processing for tables, headers, and section formatting; built a **section-aware chunking strategy** and an **LLM skeleton parser** (Claude Haiku) that reads every page, summarises each section, and constructs a full document hierarchy for semantically rich indexing.
- Identified that users map source sections to target sections, not text chunks — making conventional RAG the wrong model. Designed **agentic section-RAG** with a bookmark extraction tool: agent extracts the full document bookmark hierarchy, selects the top 3 relevant sections per query, and returns them as grounded context — significantly improving retrieval precision for long-form regulatory authoring.
- Designed a **Planner-Executor multi-agent framework** across 12 document types; built a **base agent class** with **LangSmith** observability, **Portkey** for LLM proxy and prompt versioning (managing 100s of prompts and model configs team-wide), and short-term memory. Deployed centralized knowledge as an **MCP server on AWS Agent Core**; established **LLM-as-a-judge eval** with golden responses and **ACS-framework** prompts via LangSmith; human validation checkpoints at critical stages for FDA compliance.

Clinical Query Analysis System

- **Problem:** Clinical teams had no objective framework to assess query quality — query tagging was entirely subjective. A stratified sample analysis revealed **14% redundancy**, significant enough to justify building a systematic classification pipeline.
- Built a **query classification framework** applying three independent classifiers to every query: **Query Type** (data mismatch, labelling issue, missing information, etc.), **Query Quality** (actionable vs. non-actionable — based on clarity, presence of next action, and whether a genuine issue was raised vs. user frustration), and **Redundancy Detection** against a corpus of ~1.3M historical queries using LLM classification combined with statistical analysis. Cross-dimensional breakdown by query type and business group surfaced where low-quality and redundant queries concentrated.
- Deployed as an AI validation layer on the live clinical query portal — whenever a user raises a query, the system automatically flags redundancy, assigns a quality score, and tags it with a query type in real time.

Data Science Associate Consultant | ZS Associates
Dec 2023 – Dec 2025 · New Delhi

Patient Safety Narrative Authoring

- **Problem:** Original brief: a drafting copilot for patient safety narratives. Identified the real bottleneck was **validation** — verifying every LLM statement against patient data scattered across disconnected silos. Built an **ETL pipeline** consolidating adverse event data into a **per-patient fact table** visualised as a **Gantt chart of event timelines** — giving medical writers a single source of truth for sign-off and grounding the LLM at generation time.

- Designed a **3-step generation pipeline**: (1) **Data-to-Facts generation** with code-level validation (dates, medication names, adverse events verified programmatically); (2) **Narrative Refinement module** applying FDA writing guidelines; (3) optional **Augmented narrative generation** for clinician notes and additional documents. **CoT, few-shot, and temporal reasoning prompting**; scaled to 600+ patients across 3 trials.

Auto Document Redactor

- **Problem:** Regulated clinical documents required PII redaction before external sharing — errors risked compliance violations in FDA-governed workflows. Designed a **hybrid redaction pipeline**: rule-based heuristics for deterministic patterns + **fine-tuned spaCy NER** trained on clinical PII + **chain-of-thought prompting** for context-sensitive entities — achieving **98% recall / 90% precision**. Built evaluation benchmarks for repeatable compliance testing; enabled confident pipeline iteration without regression.

Data Science Associate | ZS Associates *Nov 2021 – Dec 2023 · New Delhi*

ISR Entity Detection Pipeline

- **Problem:** Stakeholders needed to make data-driven ISR sponsorship decisions, but relevant information — drug names, dosages, treatment cycles, endpoints, inclusion/exclusion criteria — was buried in unstructured clinical protocol documents. Built an end-to-end **entity extraction pipeline** combining **spaCy NER**, custom **fine-tuned entity recognition models**, and **XGBoost classifiers** to structure protocol intelligence at scale for systematic comparison of ISR proposals.

Patient Discontinuation Prediction

- **Problem:** Client needed early identification of patients at risk of therapy discontinuation to enable timely retention interventions.
- Developed an **XGBoost model** with features engineered from claims data, patient demographics, and census datasets; used **k-fold cross-validation** for robust performance estimation and **RFE with forward/backward elimination** for feature selection; packaged as a **reusable cross-client asset** enabling client-specific model retraining across ZS engagements.

Research Intern | Inria, University of Lille *Jun – Sep 2021 · Lille, France*

- Modeled suicide attempt recidivism in mental health patients using parametric/non-parametric survival methods with geostatistical spatial autocorrelation to account for regional variation in re-attempt probability.

ML Engineer (Part-Time) | Omdena / World Resources Institute *Aug 2020 – Feb 2021 · Remote*

- CNN multi-task model on Landsat 7 & 8 satellite imagery + DHS/census data predicting district-level Asset Wealth Index across India; hosted by WRI under UN SDG 8 (team of 50).

PERSONAL PROJECTS

CalorieBot — Nutrition Tracking Agent · 2025 · GitHub

- **LangGraph** 6-phase agentic pipeline (WhatsApp bot) with **short-term + long-term memory**: meal input → parsing → parallel food search → macro scaling → logging → daily summary.
- Configured **Claude Code environment** (CLAUDE.md, skills, rules) for backend API + iOS; shipped companion app to **TestFlight**; beta-tested with real users, iterated on prompts and search scoring from their feedback.

EDUCATION

BSc (Hons) Statistics · Kirori Mal College, University of Delhi · CGPA: 8.6 *Graduated July 2021*