

AAKASH SANGANI

AI governance and auditable AI research

Mumbai, India · aakash.sangani2410@gmail.com · [linkedin.com/in/aakash-sangani-3149a11a7](https://www.linkedin.com/in/aakash-sangani-3149a11a7) · github.com/aakash2410 · aakash2410.github.io

RESEARCH INTERESTS

Auditable and accountable AI systems: inferring the revealed objectives of deployed agents from their behaviour, the reliability of the evaluators we use to make safety claims, algorithmic fairness in Indian credit markets, and conformity assessment and incident reporting under the EU AI Act and California's SB-53.

EDUCATION

B.Tech, Information Technology

2019 – 2023

D. J. Sanghvi College of Engineering, University of Mumbai

CGPA 8.90 / 10

- Merit scholarship covering 100% of tuition. Coursework: Probability & Statistics (O, 93%), Advanced Data Structures (O, 95%), Artificial Intelligence (A+, 84%), Statistical Analysis (A+, 81%), Design & Analysis of Algorithms (A+, 83%).

RESEARCH AND INDEPENDENT WORK

Inverse-Objective Regulator

github.com/aakash2410/Inverse-Objective-Regulator

- Black-box auditing system that infers a deployed agent's revealed objective from its tool-call traces via Bayesian inverse reinforcement learning, scores divergence from its declared purpose, and emits regulator-legible evidence mapped to the EU AI Act, OWASP MITRE ATLAS and NIST AI RMF.
- Open 8-agent benchmark of agentic failure modes (reward hacking, power-seeking, sycophancy, privacy leakage, safety-skipping). A deterministic open cross-encoder critic ranks the planted objective top-2 on every seed for 6 of 8 agents (mean $r = 0.55$), deterministically and with no API cost, against a frontier LLM-as-judge spanning 0.08–0.43 across agents. Sole author of the paper draft (not yet on arXiv) and the implementation.

Credit Equity Audit of India (NSSO 77th Round)

github.com/aakash2410/CreditAudit-India

- Two-stage, Heckman-style framework separating access to the credit market from allocation of formal versus informal credit. Female-headed households had ~59% of the probability of market entry of male-headed households (disparate impact 0.59), against the conventional 0.8 threshold.
- IBM AIF360 adversarial debiasing and SHAP to test for proxy discrimination, lifting disparate impact from 0.59 to 0.94 at ~73.2% accuracy. Documented the design's limits: self-reported survey data, collider bias risk, and SHAP as a model diagnostic rather than causal evidence.

The Specialist, on the Suno Sutra edge platform

github.com/aakash2410/SpecialistAISunoSutra

- Grounded offline retrieval with explicit abstention on the open-source AI device from Digital India Bhashini (MeitY) and Current AI, which runs vision, ASR, translation and TTS on-device across 22 Indic languages. Answers are grounded in approved DIKSHA content and carry a citation to chapter, section and page.
- Two gates enforce abstention: retrieval refuses below a similarity threshold before an LLM call is spent, and the grounded prompt makes the model emit a no-answer signal the pipeline converts into a refusal. multilingual-e5-small embeddings with a deterministic TF-IDF fallback keep the loop reproducible without a neural model.
- Designed the device's inference pipeline and owned the benchmarking behind it: batching and scheduling tradeoffs across three concurrent models under a hard RAM ceiling cut latency to first audio from ~8s to 2–3s and peak memory by 40%.

rag-compliance-auditor

github.com/aakash2410/MCP_RAG_Compliance

- Sole-authored package on PyPI. MCP server that audits RAG and AI pipelines against the EU AI Act, GDPR/DPD, HIPAA, SEBI/RBI and NIST AI RMF across 7 audit dimensions; RS256-signed audit reports, 8 MCP tools, CI gate with SARIF reporting.

PUBLICATIONS AND POLICY WRITING

Sangani, A., Doshi, P., & Sawant, V. "Unusual Social Media Behavior Detection Using Distributed Data Stream Mining." In *Practical Data Mining Techniques and Applications*. Auerbach Publications (CRC Press), 2022–23. Clickstream-based detection of hacked or bot-controlled accounts: per-cluster Very Fast Decision Trees on Apache Spark, with concept-drift handling.

Sangani, A. *Inverse-Objective Regulator*. Paper draft, sole author, not yet on arXiv.

Section on AI incident-reporting mechanisms for India, in *AGI Governance*, with the Office of the Principal Scientific Adviser to the Government of India, drawing on the EU AI Act and California's SB-53.

EVALUATION AND ASSESSMENT DESIGN

- Designed the evaluation framework for the **VYOMA Innovation Challenge** (₹80 lakh, Digital India Bhashini × Current AI × Kalpa Impact): jury rubric, use-case judging matrix, and the scoring thresholds applied through to the finale at IIIT-Delhi.

- Verified edge-AI performance claims submitted to the Gates Foundation; wrote the reference architecture for AI-enabled coordinated care over ABDM-connected health systems.

PROFESSIONAL EXPERIENCE

AI Consultant

Kalpa Impact, Mumbai

Digital public infrastructure and AI governance for UNDP, DPGA, Gates Foundation, Co-Develop

Feb 2026 – present

Business Technology Solutions Associate → Senior Associate

ZS Associates, Pune

Sep 2023 – Jan 2026

- Launched an agentic AI analytics product for natural-language querying of sales-performance metrics; owned incentive-compensation product logic for a Fortune 100 pharma client at 100% accuracy. Led a team of 4.

Product Analyst Intern

Bright Money, Bangalore

Jul 2022 – Feb 2023

- Built A/B testing infrastructure and SQL dashboards; launched a referral programme that saved \$120K in acquisition cost over 5 months.

SKILLS AND HONOURS

Methods: Bayesian inverse RL, LLM evaluation and red-teaming, algorithmic fairness (AIF360, SHAP), Heckman-style selection models, RAG, edge inference and benchmarking, A/B testing

Tools: Python, SQL, ONNX, sentence-transformers, TensorFlow, scikit-learn, pandas, PostgreSQL, Apache Spark, Docker, Ansible, AWS

Honours: full-tuition merit scholarship, D. J. Sanghvi College of Engineering (2019–2023); top 1% performer firm-wide at ZS Associates, promoted within 8–10 months.