Executive Summary — GenAI Engineering

Sierra Napier

May 2026

GenAI Engineering — Executive Summary

Sierra Napier | Data Scientist | AI Architect

What I Built

Three real-world NLP systems using live research, legal, and biomedical APIs.

Project	Domain	Data Source	Key Result
arXiv Research NLP	Research trend analysis	arXiv API: 450 papers (cs.LG, cs.AI, cs.CL, cs.CV, stat.ML)	Category distribution + trend identification
SCOTUS Opinion Mining	Legal text analysis	CourtListener API: 15 landmark opinions (1954–2015)	Opinion length trends, sentiment shifts, citation networks
PubMed Biomarker Analysis	Biomedical research mining	NCBI E-utilities: Immunotherapy trials + cancer biomarkers	Volcano plots, differential expression, pathway enrichment

---

Why This Matters

For Research Teams: The arXiv pipeline automates literature review. Instead of manually tracking 450+ papers, you get interactive dashboards showing which ML subfields are growing fastest.

For Legal Departments: SCOTUS opinion mining demonstrates how to extract structured insights from unstructured legal text. Same approach scales to contracts, regulations, and compliance documents.

For Biotech/Pharma: PubMed biomarker analysis shows how to process clinical literature at scale. The volcano plot and pathway analysis are directly applicable to drug discovery pipelines.

Tech Stack

Python · BERT · spaCy · TF-IDF · LLM Prompt Engineering · NCBI E-utilities · Plotly

---

Verification

All papers fetched live from arXiv API. All SCOTUS opinions from public domain CourtListener. All PubMed records via NCBI E-utilities. No synthetic abstracts.

Live Code: github.com/gosidehustlesisi/sierra-genai-engineering
Interactive Portfolio: gosidehustlesisi.github.io/sierra-applied-ml/

"GenAI isn't about generating content — it's about extracting signal from noise in specialized domains."