

Tracking Evolving Skill Demands in Data & AI Jobs with LLMs

Fanchao Lin
STAT418 Final Project Proposal

WEBSRAPE: Extracting Data

- Scraped job postings from Indeed using BeautifulSoup (static HTML): Extracted job title, location, company, posting date, salary, and full description for Data/AI-related roles and saved them to a .csv file.
- Sample GPT-4 API output for keyword summarization of the above job description:

Data Scientist

UCLA Health 

(part of University of California Los Angeles) • 4.0 ★

Los Angeles, CA 90095

\$102,500 - \$227,700 a year

You must create an Indeed account before continuing to the company website to apply

Apply now 



```
{
  "job_title": "Data Scientist",
  "company": "UCLA Health",
  "extracted_insights": [
    "emerging_ai_frameworks_or_tooling": [
      "LLMOps (Large Language Model Operations)",
      "Generative AI Applications",
      "MLOps (Machine Learning Operations)",
      "Responsible AI Governance",
      "Bias & Fairness Testing"
    ],
    "industry_contextualized_use_cases": [
      "Clinical AI Optimization",
      "Operational Process Enhancement",
      "Financial Outcome Forecasting",
      "Equity and Ethics in Healthcare AI"
    ],
    "company_industry": "Healthcare (Academic Medical Center)"
  ]
}
```

EXPLORATORY ANALYSIS

Top 10 Hard skills:

Top 10 industry applications:

Top 10 hiring industries:

Rank	Hard_skill
1	LLMops
2	MLOps
3	Generative AI
4	Responsible AI & Bias Testing
5	Prompt Engineering
6	Knowledge Graphs / Ontologies
7	Vector Search / Embedding Stores
8	Reinforcement Learning / RLHF
9	Agentic AI Frameworks
10	Advanced NLP / NLU Techniques
Rank	Application
1	Healthcare optimization
2	Financial fraud detection
3	Retail personalization
4	AI-assisted drug discovery
5	Customer service automation
6	Supply chain forecasting
7	Document automation
8	Education technology
9	Autonomous operations
10	Sustainability & energy
Rank	Industry
1	Healthcare & Biotech
2	Finance & Insurance
3	Technology (Big Tech & AI startups)
4	Retail & E-commerce
5	Manufacturing & Industrial Automation
6	Energy & Utilities
7	Government & Public Sector
8	Education Technology (EdTech)
9	Logistics & supply chain
10	Media & Entertainment

PROJECT PROPOSAL

- Expand data collection to include additional sources (e.g., LinkedIn).
- Use Shiny to integrate the entire workflow: web scraping, GPT API calls, summarization, and result visualization; deploy the application using Docker.
- Answers key questions of job market trends: What advanced skills and experiences are currently in demand? Which industries are actively hiring for these roles? How is the number of job postings evolving over time?