
DHUM 25A43: Investigating with AI

Fall 2025 | Sciences Po

Course Description

This course bridges traditional social science research with contemporary AI capabilities, empowering students to gather, transform, and analyze textual data at scale. Students become active investigators who extract meaningful patterns from digital information using AI tools, APIs, natural language processing, and data visualization.

Course Philosophy

This is a hands-on course emphasizing learning by doing. Students work collaboratively on real-world data projects, building AI literacy through direct experience with cutting-edge tools.

Learning Outcomes

By the end of this course, students will be able to:

- Collect and process data from web APIs and online sources
 - Apply NLP techniques to analyze textual data at scale
 - Build AI-powered applications using LLMs and agents
 - Create data visualizations that communicate insights effectively
 - Deploy interactive web applications for data analysis
 - Critically evaluate AI tools and their societal implications
-

Course Schedule (12 sessions, 2 hours each)

Session 1: Welcome & Introduction

- Course scope, logistics, and evaluation

- State of AI: LLMs, benchmarks, and capabilities
- Hands-on: Data analysis with Google Colab and Gemini

Session 2: Data Foundations

- Open source vs. closed source AI models
- Data sources and datasets (Kaggle, HuggingFace, government data)
- Data formats: JSON, CSV
- The Pandas library for data manipulation

Session 3: Web & APIs

- How the web works: HTML, URLs, REST protocol
- Building datasets from APIs
- Hands-on: Wikipedia API, NYT API

Session 4: Classic NLP

- Natural Language Processing fundamentals
- Tokens, NER (Named Entity Recognition), POS (Part of Speech)
- Text classification: sentiment, topic modeling
- Hands-on: Spacy library

Session 5: Modern NLP

- Word embeddings and semantic similarity
- LLMs and contextual understanding
- RAG (Retrieval Augmented Generation)
- Hands-on: Building embeddings-based search

Session 6: Building Web Applications

- Web technologies: HTML, CSS, JavaScript
- Streamlit for rapid prototyping
- Deploying applications to GitHub

Session 7: From LLMs to Agents

- Augmented LLMs: tools, memory, planning
- Agent frameworks: multi-agent systems
- Function calling and Model Context Protocol (MCP, Claude Skills)

Session 8: Data Visualization & Inference

- Principles of effective data visualization
- Data narrative and storytelling

- Fast inference via APIs: Groq, OpenRouter
- Hands-on: Comparing models, LLM-as-judge

Session 9: Networks & Web Scraping

- Network analysis fundamentals
- Web scraping techniques and ethics
- Building network visualizations
- Hands-on: Extracting and analyzing network data

Session 10: Machine Learning & Deep Learning

- Supervised learning: classification and regression
- Deep learning fundamentals
- Introduction to reinforcement learning
- Hands-on: Training models on project datasets

Session 11: Student Project Workshop

- Project troubleshooting and refinement
- Advanced techniques based on student needs
- Final presentation preparation

Session 12: Final Presentations

- Group project presentations to class and experts
- Peer feedback and discussion

Evaluation

Group Project (80%) - Website or interactive report showcasing data investigation - Public presentation to class and invited experts - Evaluation criteria: data quality, analysis depth, AI literacy demonstration, presentation clarity

Individual Reflection (20%) - Critical reflection on working with AI tools - Personal propositions for AI use in social sciences - 3-5 pages, due at end of semester

Required Tools & Resources

- **Google Colab** for Python programming and data analysis
- **Course website:** skatai.com/inwai
- **GitHub account**
- **Instructor:** Alexis Perrier, alexis.perrier@sciencespo.fr

