

How I Use AI in My Academic Research

Xiaoxuan (Cleo) Xu

PhD Candidate in Information System & Analytics

“You don’t have to become a technical expert before asking the Research Question you care about.”

What I study

Online Communities and Platforms

Group dynamics, content moderation, and platform governance.

Multimodal Digital Discourse

Podcasts, audio narratives, text, and computational approaches to culture and storytelling.

Computational Social Science

NLP, LLMs, embeddings, causal inference, econometrics, and network analysis.

My work studies how individuals and online communities organize, communicate, and create cultural meaning in platform-mediated social and economic life.



Which AI I use for research

- **Conversational / Interactive AI** (Chatbot, Code, and Cowork)



- **Programmable Language AI** (APIs or local LLMs)

Frontier models

Start with GPT-5.6 Sol for complex reasoning and coding, choose GPT-5.6 Terra to balance intelligence and cost, or use GPT-5.6 Luna for cost-sensitive, high-volume workloads.

[View all](#)[Compare models](#)

GPT-5.6 Sol

Frontier model for complex professional work



GPT-5.6 Terra

GPT-5.6 model that balances intelligence and cost



GPT-5.6 Luna

GPT-5.6 model optimized for cost-sensitive workloads



Llama 3.1 70B

70B Parameters



Mixtral 8x7B

8x7B Parameters



Ollama



Qwen2.5 72B

72B Parameters



Yi 1.5 34B

34B Parameters



Gemma 2 27B

27B Parameters



- **Task-specific ML** (Pretrained models from open-sourced communities)



Hugging Face



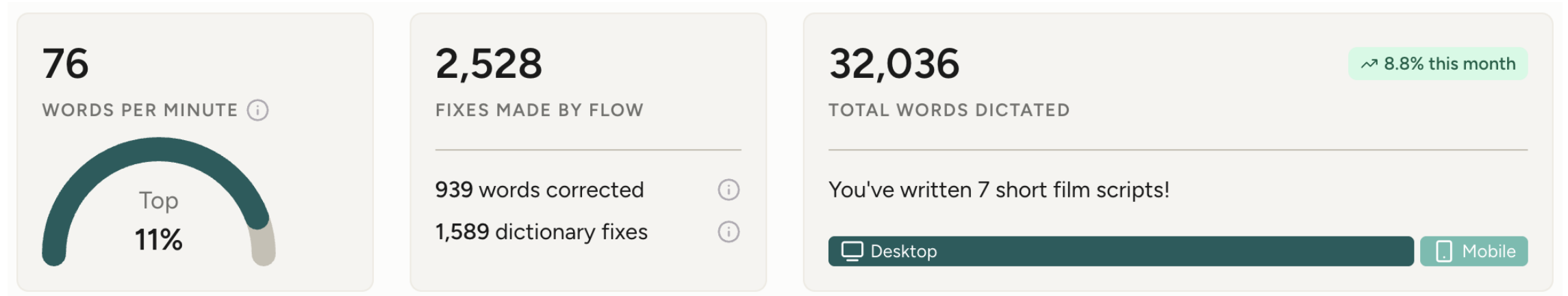
GitHub



1 Interactive AI as everyday research partner

- Brainstorming and RQ development
- Coding and debugging
- Planning and project management
- Second screening and consistency checks

1. Use audio-to-text AI to capture thoughts quickly



1 Interactive AI as everyday research partner

2. Make the project auditable

- Use a clear folder structure
- Add README and meta files
- Record versions, decisions, and outputs

AI create faster than we can remember.

```
msrb\  
├── README.md # map of the project: what lives where, how to run what  
├── PROJECT_LOG.md # append-only ledger – one dated entry per work session  
├── DECISIONS.md # numbered decision records (D-001, D-002, ...)  
├── STATUS.md # current-state snapshot (the good habit STATUS.md started)  
├── DATA_PROVENANCE.md # where every input came from, licenses, integrity checks  
├── .gitignore  
├── docs\  
│   ├── briefs\  
│   │   └── # County Subreddit Data Brief, MSRB Data Brief (+ figures)  
│   ├── design\  
│   │   └── # research_design.md, methods doc (canonical .md; .docx = export)  
│   ├── measurement\  
│   │   ├── idv\  
│   │   │   └── # rubric, anchor bank spec, measurement plan (one canonical each)  
│   │   └── (pdi|uai|mas|lto|ivr) # created as each dimension starts  
│   └── writing\  
│       └── # workshop abstract, benchmark memo, paper drafts  
├── src\  
│   ├── s0_extraction\  
│   │   └── # phase0_geo_county_extract.py, select_torrent_files.py  
│   ├── s0_panels\  
│   │   └── # phase0_build_panels.py (→ panels_v2 lives here too)  
│   ├── s1_corpus\  
│   │   └── # phase1_build_corpus.py  
│   ├── s4_retrieval\  
│   │   └── # phase4_retrieval_ic.py + requirements_phase4.txt  
│   ├── s5_annotation\  
│   │   └── # annotation tool + sampling scripts (future)  
│   ├── msrb\  
│   │   └── # MSRB cleaning & outcome construction (future)  
│   └── common\  
│       └── # config.py (paths per machine), runlog.py helper  
├── data\  
│   ├── raw\  
│   │   ├── msrb\  
│   │   │   └── # msrb_all\ + variables PDF (immutable)  
│   │   └── bots\  
│   │       └── # bot_list_a.csv, bot_list_b.csv  
│   └── reference\  
│       ├── # county_subreddit_mapping, cusip_issuer_cnty_map,  
│       └── # cusip6_to_geo_subreddit, geo_subs_1210.xlsx, msrb_cusip6.csv  
├── reddit\  
│   ├── county_monthly\  
│   │   └── # the 566 county folders (renamed from county_monthly_data)  
│   ├── metadata\  
│   │   └── # canonical monthly metadata + coverage pivots  
│   ├── extraction_qa\  
│   │   └── # ← coverage_old, renamed to say what it is  
│   ├── corpus_panel_A\  
│   │   └── # sentences\, diagnostics, author_counts, build logs  
│   └── panels\  
│       └── # panel_A/B/C.csv, overlap (+ future panels_v2)  
└── outputs\
```

2 Programmable language AI scales text analysis

API

- Fast to start
- Hosted, strong frontier models
- Pay per use

Local LLM

- More control over model and data
- Keep sensitive data local
- Requires setup and computing resources

Provider / platform	Current models / offering	Cost to try	Live docs
OpenAI	GPT-5.6 Sol, Terra, Luna	Paid; per token	Models + pricing
Anthropic	Claude Fable 5, Opus 5, Sonnet 5, Haiku 4.5	Paid; per token	Models + pricing
Google	Gemini 3.6 Flash, 3.5 Flash-Lite	AI Studio free; API free/paid tiers	Models + pricing
Groq	Fast inference: GPT-OSS, Qwen, Whisper	Free tier; paid scaling	Models + pricing
AWS Bedrock	Managed access to 100+ model options	Pay per use	Models + pricing
Ollama	Local/open models plus optional cloud	Local free; cloud free/paid	Models + pricing



Same task + same prompt (unstructured)

THE SHARED PROMPT — A SMALL FINANCIAL-ANALYSIS TASK

Act as a financial analyst. Based on the following metrics, provide a one-sentence conclusion starting with 'Signal: [Good/Bad/Neutral]' followed by the short reason why.

Net sales increased 14% to \$213.4 billion in the fourth quarter, compared with \$187.8 billion in fourth quarter 2024. Operating income increased to \$25.0 billion ... Net income increased to \$21.2 billion, or \$1.95 per diluted share ...

OpenAI

Anthropic

Google

Ollama (local, free)

Compare all four

OpenAI · gpt-4o-mini

Signal: Good, as all key financial metrics—net sales, operating income, and net income—show significant improvements year-over-year, indicating a strong performance.

Anthropic · claude-haiku-4-5

Signal: ****Good**** – Strong double-digit revenue growth of 14% combined with operating income increasing 18% and net income rising 6% with improved EPS demonstrates robust profitability expansion and operational efficiency gains.

Google · gemini-2.5-flash

Signal: Good, driven by substantial increases in net sales, operating income, and net income across the board.

Ollama (local, free) · llama3

Signal: Good

The strong increase in net sales, operating income, and net income indicates a positive trend for the company, suggesting that its business is performing well and generating significant revenue and profitability growth.

LLM for info extraction (structured)

2.1 Extraction & NER

2.2 Zero-shot classification

2.3 Reproducibility

2.4 Majority voting

2.5 Topic labeling

2.6 Aspect sentiment

2.7 Embeddings

One prompt replaces the whole NLTK pipeline

Prior lectures: tokenization → POS → NER → regex, all hand-rolled. With an LLM: describe the schema you want, in words.

INPUT — A RAW JOB POST

Avery Dennison is looking for a Materials Scientist - Machine Learning.

Requirements: Ph.D. in Polymer Science, Chemical Engineering, or Physics. Experience with Python, TensorFlow, PyTorch, C++.

Familiarity with Molecular Dynamics and Finite Element Analysis. Location: Mentor, OH. Contact: xxx@stevens.edu

Extract entities (llama3, JSON mode)

```
{"company": "Avery Dennison",  
  "job_title": "Materials Scientist - Machine Learning",  
  "skills": ["Python", "TensorFlow", "PyTorch", "C++"],  
  "location": "Mentor, OH",  
  "email": "xxx@stevens.edu"}
```

Podcast project task: extract guest speakers' info

A	B	C	D	E	F	G	H
Filename	Episode Title	Cleaned Description	Publication Date	Guest Name	Company Name	Role	Company Website
bswra_1	Becoming "The One Percent" With	In 2008 Scott Neal hit a	7/7/23	Scott Neal	Taylor Estridge	Founder	https://taylorestridge.com
bswra_2	How Robotics Can Take Your Op	As a young boy I was fasc	7/27/23	Timothy Tanksley	RichTech Robotics	Founder	https://www.richtechrobotics.com/
bswra_3	Discussing A New Approach To F	Born to teenage parents	8/2/23	James Webb	Orangetheory Fitness	Franchise Owner	https://www.orangetheory.com/en-us
bswra_3	Discussing A New Approach To F	Born to teenage parents	8/2/23	James Webb	ScentHound		https://www.scenthound.com/
bswra_4	Entrepreneurial Lessons From a	What does a lifelong ha	8/18/23	Jason Hamacher	Lost Origins Gallery		https://lostorigns.gallery/
bswra_5	Unveiling the Sales Surge: Ecom	Kristine Huckell (https://www.ecomaids.com)	8/24/23	Kristine Huckell	Ecomaids		https://www.ecomaids.com/claremont-ca/
bswra_6	Dialing Up Profits: How the Right	Join us in this episode a	8/29/23	Heather Parker	Hutchbug Solutions	Founder	https://www.hutchbug.com/
bswra_7	Revolutionizing Sales: Unveiling	In this eye-opening epis	8/29/23	David Newman		Founder	https://doitmarketing.com/selling/
bswra_9	The Franchise Journey with Bisc	On this episode I speak	9/22/23	Chad Coulter	Biscuit Belly	Founder	https://biscuitbellyfranchise.com/
bswra_10	From Passion to Profit: Jesse Key	In this captivating episo	10/10/23	Jesse Keyser	Multi-Unit Franchising Conference		https://www.multiunitfranchisingconference.com/
bswra_11	Sandler Training Uncovered With	In this episode with San	10/11/23	Algot Sorensen	Sandler Training		https://www.sandler.com/
bswra_12	Elevating the Roofing Industry: A	In this episode, we sit d	10/12/23	Mike Harvey	The Roof Resource	Founder	https://theroofresource.com/
bswra_15	Unlocking the Future of Immigra	Join us on an enlighteni	10/25/23	Leonardo Shapiro	Not Alien	Founder and CEO	https://notalien.com/
bswra_16	Flame-Kissed Success: Unveilin	Unlock the door to entre	11/9/23	Joseph Becker	Woody's BBQ		https://woodys.com/franchise/
bswra_17	From Vision to Expansion: Unvei	In this episode, join me	11/15/23	Rick Robinson	Franchise Genesis	Founder	https://www.franchisegenesis.com/
bswra_19	Behind the Valuation: Navigatin	Join us in this insightfu	12/14/23	Don Preston	Home Appraisal Solutions		https://www.homeappraisalsolutions.com/
bswra_20	From Dreams to Reality: Unlock	Join us for an insightfu	12/19/23	Kim Daly	FranChoice		https://www.franchoice.com/
bswra_21	A Deep Dive into the World of Fr	Embark on a fascinating	1/15/24	Irving Chung	FranChoice		https://www.franchoice.com/
bswra_23	James Hilovsky And His Game Pl	On this episode, expert	3/11/24	James Hilovsky	The Fran Dream		https://www.thefrandream.com/
bswra_27	Ep 107 Painting a Picture of Succ	Join us on episode #107	4/22/24	Walter Nunez	A&P Painting and Flooring		https://appaintingandflooring.com/
bswra_28	Ep 106 Achieving Scale with a H	Welcome to another ep	4/22/24	Aaron Santas	RoofCoach	Founder	https://roofcoach.net/
bswra_29	Ep 108 Pipes & Dreams: Steve P	In this inspiring episode	4/23/24	Steve Paley	1800PlumberAir		https://1800plumberfranchise.com/
bswra_30	Ep 109 How To Leverage Custom	In episode #109, join us	4/25/24	Sebastian Aguirre	NiceJob		https://get.nicejob.com/
bswra_31	Ep 110 Enhancing Your Home wi	Join us on episode 110 a	4/26/24	Dustin Ingle	Insulation Commandos	CEO	https://insulationcommandos.com/
bswra_32	Ep 111 Creating Email Copy Tha	Join us on episode #111	4/30/24	Ian Stanley	Email Game Changers	Founder	https://www.emailgamechangers.com/
bswra_33	Ep 112 Tiny Spaces, Big Dreams	Join us as we delve into	5/1/24	Colton Paulhus	Anchored Tiny Homes	Founder	https://anchoredtinychomes.com/

Podcast project task: label RST relation (human vs llm)

type: all

show: all

gold label: all

marks: all

Export fixed CSV

268 pairs

gold: elaboration

gemma3_12b_pv1: elaboration

llama3.1_8b_pv1: elaboration

qwen3_32b_pv1: continuation

qwen3_8b_pv1: elaboration

☐ hard (audio/seg issue)

☐ prompt example (models wrong)

R0006

100me_260218 | within | pause 0.04s

They did have a good leadership team. And they had leaders

who were selling

and were leading,

which is a really bad thing. So it took a while.

gold: elaboration

gemma3_12b_pv1: elaboration

llama3.1_8b_pv1: elaboration

qwen3_32b_pv1: continuation

qwen3_8b_pv1: continuation

☐ hard (audio/seg issue)

☐ prompt example (models wrong)

note: parallel

R0007

100me_260218 | within | pause 0.3s

And they had leaders who were selling

and were leading,

which is a really bad thing.

So it took a while. I mean, it took a couple of three years of work to get all those pieces in place

gold: evaluation

gemma3_12b_pv1: evaluation

llama3.1_8b_pv1: evaluation

qwen3_32b_pv1: evaluation

qwen3_8b_pv1: evaluation

☐ hard (audio/seg issue)

☐ prompt example (models wrong)

R0008

100me_260218 | within | pause 0.821s

who were selling and were leading,

which is a really bad thing.

So it took a while.

I mean, it took a couple of three years of work to get all those pieces in place because there was no way

3 Task-specific ML

Pretrained models and open-source implementations, mostly from GitHub and HuggingFace.

The screenshot shows the Hugging Face website interface. At the top, there's a search bar and navigation links for Models, Datasets, Spaces, Buckets, and Docs. The 'Main' tab is selected, showing filters for Tasks (Text Generation, Any-to-Any, Image-Text-to-Text, Image-to-Image, Text-to-Image, Text-to-Video, Text-to-Speech), Parameters (ranging from <1B to >500B), and Libraries (PyTorch, TensorFlow, JAX, Transformers, Diffusers, GGUF). A list of models is displayed on the right, including moonshotai/Kimi-K3, baidu/Unlimited-OCR, poolside/Laguna-S-2.1, DavidAU/Qwen3.6-27B-Fable-Fusion-711-Uncensored-Heretic-NM-D, and upstage/Solar-Open2-250B.

A smaller specialized model may be cheaper, faster, more consistent, easier to fine-tune, and easier to evaluate.

Podcast project task: speech-to-text; text-classification

Natural Language Processing

Text Classification

Token Classification

Table Question Answering

Question Answering

Zero-Shot Classification

Translation

Summarization

Feature Extraction

Text Generation

Fill-Mask

Sentence Similarity

Text Ranking

Audio

Text-to-Speech

Text-to-Audio

Automatic Speech Recognition

Audio-to-Audio

Audio Classification

Voice Activity Detection

Qwen/Qwen3-ASR-1.7B

Automatic Speech Recognition • 2B • Updated Jan 29 • 1.82M • •

openai/whisper-large-v3-turbo

Automatic Speech Recognition • 0.8B • Updated Oct 4, 2024 • 8.41M •

CohereLabs/cohere-transcribe-arabic-07-2026

Automatic Speech Recognition • 2B • Updated 15 days ago • 39.4k •

Qwen/Qwen3-ASR-0.6B-hf

Automatic Speech Recognition • 0.8B • Updated 6 days ago • 137k •

ai-sage/GigaAM-Multilingual

Automatic Speech Recognition • Updated 14 days ago • 5.02k •

Why not use one LLM for everything?

Research concerns

- Cost and efficiency
- Reliability and consistency
- Task fit and privacy

Research needs

- A fixed model version as delivery/artifact
- Fine-tuning on human-coded data
- Held-out evaluation
- Precision, recall, and error analysis



Using more AI creates more responsibility

Record:

- document prompts, model versions, data selection, and thresholds
- report errors, uncertainty, and disagreement
- preserve processes and files so others can replicate/audit the workflow

Validate:

- build human-annotated gold-set
- evaluate AI outputs

Research is both a skill set and a mindset

Industry-facing projects

- Often start with a specific goal or decision
- Work in a conditioned setting
- Optimize for delivery, deadline, or performance

Academic research

- Often starts with a more self-directed question
- Still depends on careful sampling and testbeds
- Should have some generalizability



Resources

Learn:

- Using AI in academic research — Stanford Library guide: https://guides.library.stanford.edu/ai_research
- Free ML basics — Kaggle Learn, Google ML Crash Course, 3Blue1Brown neural networks
- How to write better prompts - <https://www.promptingguide.ai/>

Compare models:

- <https://arena.ai/> — blind-compare chatbots, community leaderboard
- <https://whatllm.org/best-ollama-models> — local models for your laptop, run them with ollama.com

Open-source:

- <https://github.com/> — source code and research pipelines
- <https://huggingface.co/> — models, datasets, and free courses (hf.co/learn)





Thank you!

Xiaoxuan (Cleo) Xu

xxu62@stevens.edu

<https://cleo-xu.github.io/>