AI and Data Twister: Studying the St Louis Tornado

Adam Wilcox, PhD

Institute for Informatics, Data Science & Biostatistics

Professor of Medicine

Washington University School of Medicine

ChatGPT User Statistics

- Fastest adopted technology ever
  - In just 5 days, > 1M users
  - 100M users in 2 months
- Wide, persistent adoption
  - Top 10 on most visited websites (>5B visits/month)
  - 800M weekly active users
  - 15% of users are from US
- User base
  - 45% of users are under age 25
  - 15% of 18-29-year-olds have used it to generate text

Generative AI Use

- 56% of US workers are using Gen AI for work tasks
  - 31% using on a regular basis
- Estimates of 30-40% productivity gains using Gen AI
- Common types of use:
  - Drafting written content (68%)
  - Brainstorming ideas (60%)
  - Conducting background research (50%)
  - Analyzing data and making forecasts (19%)
  - Generating/debugging code (11%)
  - Image recognition & generation (7%)

LLM Models

ChatGPT (OpenAI)	Released, Nov 2022 – Great with reasoning via reinforcement learning – Complex reasoning and multi-step instructions – Robotic writing style, lots of repeating of jargon
Claude (Anthropic)	First hybrid reasoning model (general and reasoning tasks) – Use of language is more natural; more language styles – Not great at reasoning – High rate of hallucinations
Gemini (Google)	Largest context window that can handle more information in a conversation –Integrates with Google tools – Innovations around different models –Deep Research (search and analysis)
Copilot (Microsoft)	Similar to ChatGPT, but can be less detailed in responses; smaller context window – Responses include links to websites – Integrates directly with Microsoft ecosystem
Meta AI (Meta)	Variety of content generation – Can animate images – Integrates into social media apps
Grok (xAI)	More real-time data – Better in math and science answers, including coding
DeepSeek	Open source– Running in China, but also runs on other platforms – Can be installed locally; may be legal issues for use in US
Perplexity AI	LLM-based pipeline for web search – Interface for asking questions, with answers provided by web searches and “interpretation” (issues with copyright infringement) – Uses OpenAI’s GPT – Other LLMs have followed “deep research” capabilities

What Factors to Consider with LLMs?

- Cost effectiveness
  - Many free
    - Institutional licenses
  - Monthly fees ~$20
  - As high as $200
- Data privacy
  - Use of PHI
  - Use of your information
- Familiarity
- Workflow integration
- Content
- Strengths
- Weaknesses

All of these factors are moving targets

Using AI in Research Writing

- Benefits
  - Time savings
  - Managing data
  - Quality of writing
  - Editing
- Risks
  - Lack of context
  - Inaccuracies or bias
  - Plagiarism

Am J Cancer Res 2023;13(4):1148-1154
www.ajcr.us /ISSN:2156-6976/ajcr0150104

Review Article

The role of ChatGPT in scientific communication: writing better scientific review articles

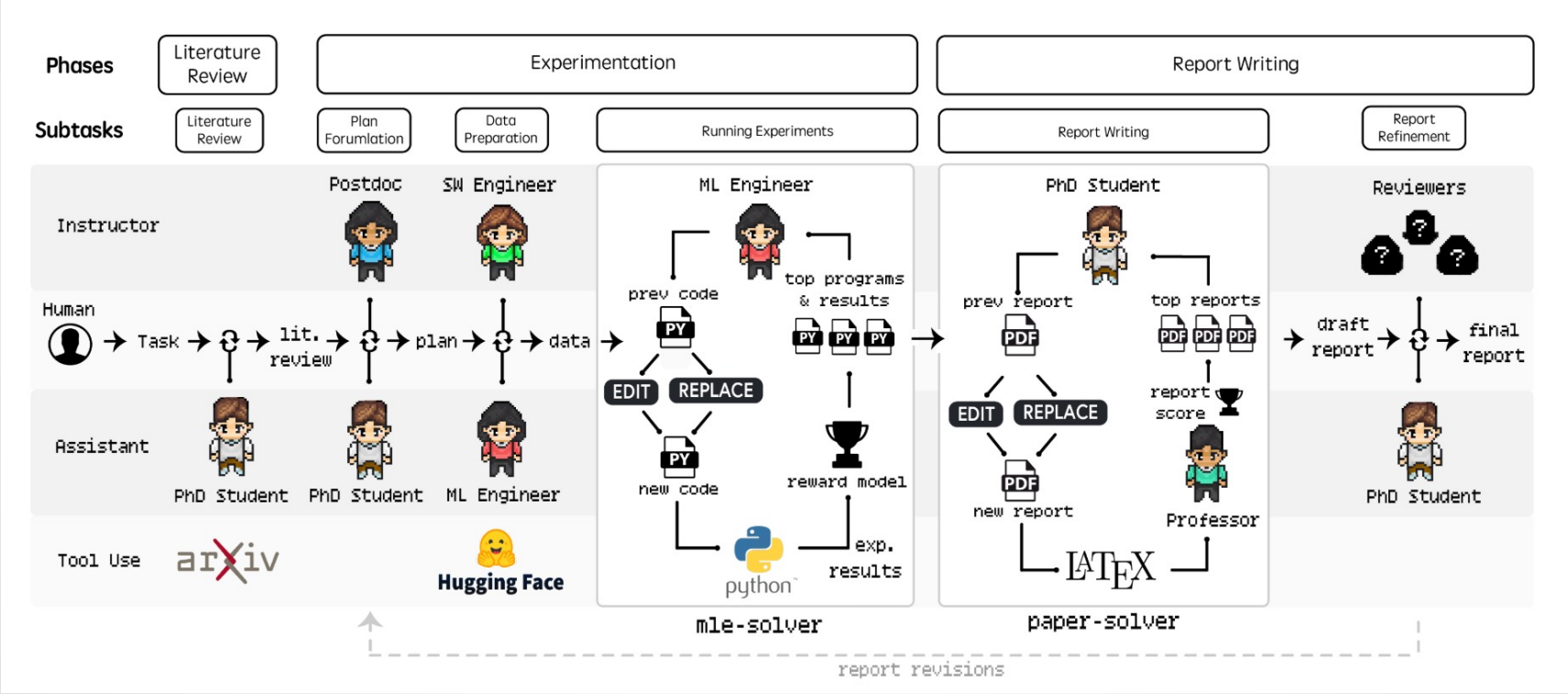
Jingshan Huang[1], Ming Tan[2]

[1]*School of Computing and College of Medicine, University of South Alabama, Mobile, AL, USA;* [2]*Institute of Biochemistry and Molecular Biology, Institute of Biomedical Sciences, and Research Center for Cancer Biology, China Medical University, Taichung, Taiwan*

Received March 13, 2023; Accepted March 23, 2023; Epub April 15, 2023; Published April 30, 2023

Abstract: Artificial intelligence tools represent an exciting opportunity for scientists to streamline their research and write impactful articles. Using artificial intelligence tools like ChatGPT can greatly improve writing review articles for scientists, by enhancing efficiency and quality. ChatGPT speeds up writing, develops outlines, adds details, and helps improve writing style. However, ChatGPT's limitations must be kept in mind, and generated text must be reviewed and edited to avoid plagiarism and fabrication. Despite these limitations, ChatGPT is a powerful tool that allows scientists to focus on analyzing and interpreting literature reviews. Embracing these tools can help scientists produce meaningful research in a more efficient and effective manner, however caution must be taken and unchecked use of ChatGPT in writing should be avoided.

AI as a Research Assistant

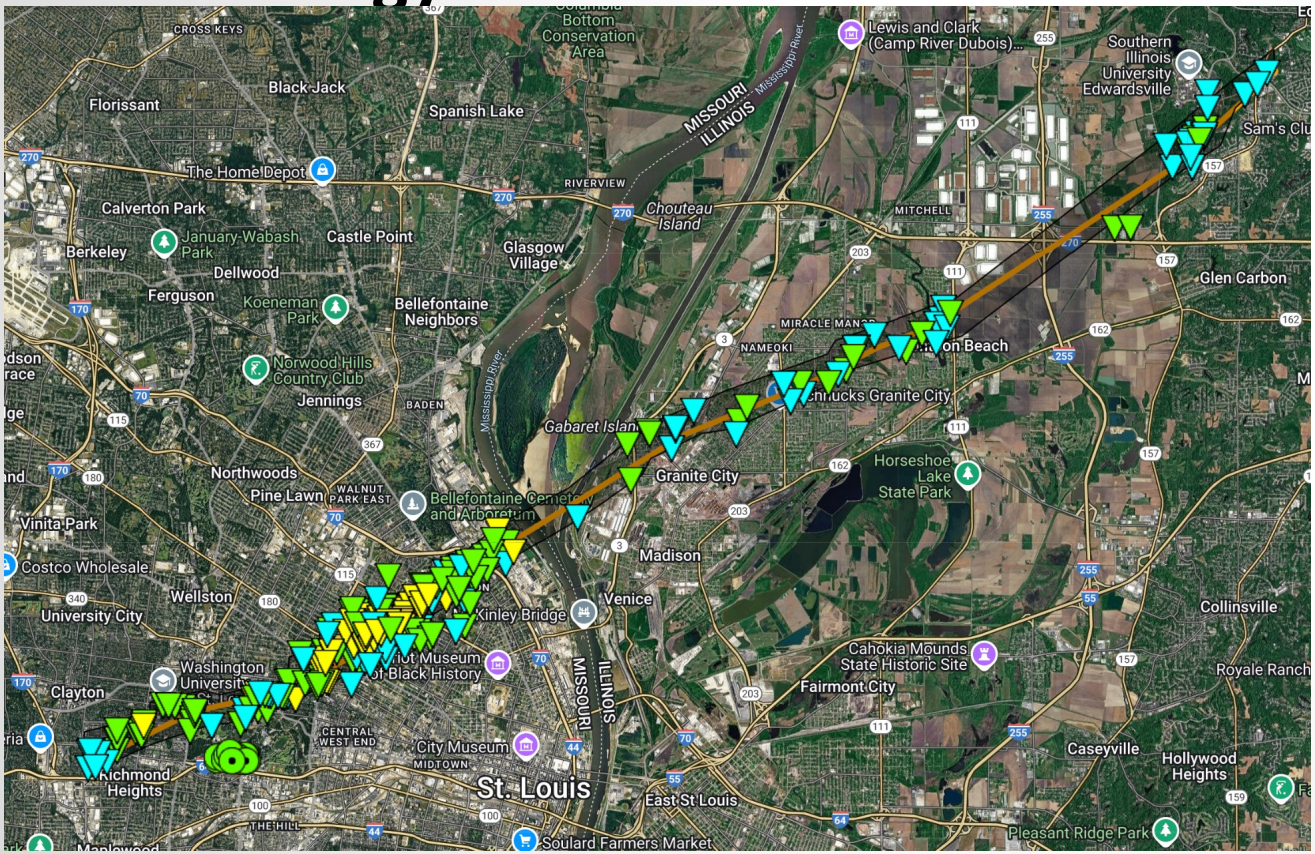


Quality of AI Output

- Novice worker: 31%
- Experienced worker: 45%
- Expert worker: 10%

Use Case: May 16 St Louis Tornado

Methodology



Start Lat/Long: 38.62966, -90.33491
End Lat/Long: 38.79989, -89.97384

Yellow line: path of tornado

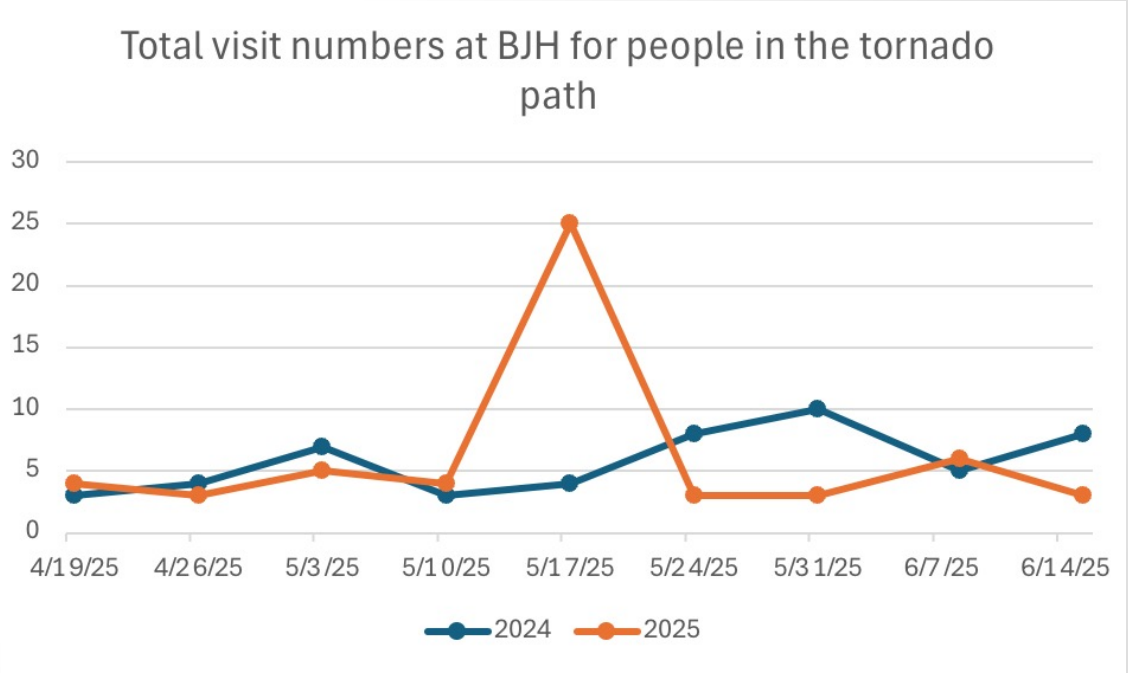
- : EF0 (65-86 MPH)
- : EF1 (86-110 MPH)
- : EF2 (111-135 MPH)
- : EF3 (136-165 MPH)

*Tornado severity determined by NWS Enhanced Fujita Scale standards

**Tornado points provided by Chief Meteorologist Brad Panovich of WCNC-TV

Results

- Total visit counts at BJH increased the weekend of the tornado in 2025. However, the same weekend in 2024 does not show this increase, leading to the conclusion that this increase in hospital visits was predominantly tornado related.



Tornado Analysis Tasks

- Identifying from notes whether visit was tornado-related
- Identifying which patients were within tornado path
- Determining payor groups to infer patient socio-economic status
- Analyzing visit types within and without tornado area
- Summarizing related literature about natural disasters and public health effects
- Writing a research report

NOTE: All data tasks will use simulated (not real) patient records

Task 1: Clinical Note Processing

Goal: Use AI to extract concepts from clinical notes that indicate whether this visit was tornado-related

- In the Task 1 folder, there are 20 emergency department notes
- First, use www.random.org to choose 10 notes to use
  - First 5 will be to read to understand note structure
  - Second 5 will be to process
- Write a prompt to interpret whether each note indicates the visit was tornado-related or not
  - Determine correctness
- Summarize results for the group

Task 2: Geospatial Analysis

Goal: Identify which patients lived within the tornado path

- In the Task 2 folder, there is a file with patient records from the Emergency Department (Task2_Data.csv)
  - Patient ID, Date, Address, Latitude, Longitude
- There is also a file that contains a link to the tornado map (Tornado_Path_Map.txt)
- Choose 3 records and from the data file, and determine if they are in the tornado path.
- Use AI to write code (R or Python) to determine whether each address is within the tornado path or not

Task 3: Classifying Payors

Goal: Identify payor classes (esp. Medicaid)

- In the Task 3 folder, there is a file with patient records (Task3_Data.csv)
  - Patient ID, Date, Address, Latitude, Longitude, Tornado_Path?, Payor
- Use AI to analyze the data and give the frequency for each payor in the data
- Identify which payors are Medicaid, Commercial, or Medicare.

Task 4: Analyzing Data

Goal: Analyze data to compare conditions within and without the tornado path

- In the Task 4 folder, there is a file with patient records (Task4_Data.csv)
  - Patient_ID, Date, Tornado_Path?, Payor, Primary_Dx, Dx_Type
- Use AI or another tool to analyze the data, comparing the Dx_Type for records within and without the tornado path
- There is also another file with records for the full month (May2025.csv). Analyze these data with a question of interest for you.

Task 4: Analyzing Data (cont.)

Goal: Review maps of tornado data

- In the Task 4 folder, there are two HTML files that show maps of the visits for the:
  - May 9 weekend (tornado_patients_map_A)
  - May 16 weekend (tornado_patients_map_B)
- Review these maps in a web browser

Task 5: Review Background Papers

Goal: Review published papers related to natural disasters and public health data studies

- In the Task 5 folder, there are 4 papers from recent JAMA issues related to natural disasters and public health
  - JAMA_Aung_2025.pdf, JAMA_Juarez_2025.pdf, JAMA_Paglino_2025.pdf,
  - JAMA_Purtle_2025.pdf
- Use AI or another tool to summarize each article, and identify similarities and differences among them
- Use AI to write a PubMed query for similar articles

Task 6: Writing a Research Report

Goal: Write a research report summarizing findings from the St Louis Tornado analysis

- Using your results from the previous tasks, use AI to write a 3,000-word research report about your findings.
  - Include Introduction/Background, Methods, Results, Discussion, Conclusion
  - You can either write the sections individually (in whatever order you choose), or in one very long prompt
    - Long prompts are best authored with a word processor and pasted into the AI prompt