The Future is Here...Almost: Evaluating the Capabilities of AI in Mental Health Surveillance

Darrah Edwards, MPH

August 06, 2026

Disclaimer

The views expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy of the Department of War, Defense Health Agency, or the U.S. Government.

The mention of any non-federal entity and/or its products is for informational purposes only, and is not to be construed or interpreted, in any manner, as federal endorsement of that non-federal entity or its products.

Affiliation: Battelle Memorial Institute, supporting the U.S. Navy and Marine Corps Force Health Protection Command, Portsmouth, Virginia, U.S., or Defense Centers for Public Health – Portsmouth, Virginia, U.S.

Funding Statement: This work was funded in part through Battelle Memorial Institute's contract with the Information Analysis Center Multiple Award Contract (IAC MAC) No. FA807518D0005-FA807523F0016: Ongoing Force Health Protection (FHP) Analysis, Assessment, and Evaluation for Navy and Marine Corps Force Health Protection Command (NMCFHPC). DE performed this work as employees of Battelle Memorial Institute.

Acknowledgment of Support and Disclaimer: This material is based upon work supported by the DoD Information Analysis Center Program Management Office (DoD IAC PMO) and sponsored by the Defense Technical Information Center (DTIC) and the Navy and Marine Corps Force Health Protection Command (NMCFHPC) under Contract No. FA807518D0005-FA807523F0016.

DISTRIBUTION A. Approved for Public Release, Distribution Unlimited. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Navy and Marine Corps Force Health Protection Command (NMCFHPC), the 774th Enterprise Sourcing Squadron (774 ESS), the Air Force Installation Contracting Center (AFICC), the Defense Technical Information Center (DTIC), the DoD Information Analysis Center Program Management Office (DoD IAC PMO), or of the institutions and companies affiliated with the authors.

Learning Objectives

- Analyze the applications of AskSage in enhancing data analysis processes.
- Evaluate the current limitations of AskSage in performing accurate and independent data analysis.
- Assess the constraints associated with resource allocation.

Case Study

- Original ask: Evaluate the relationship between Armed Forces Qualification Test (AFQT) categories and mental health conditions.
- Analyzed active-component Sailors stationed at Great Lakes Recruit Training between March and April 2024.
- Flagged mental health diagnoses from four datasets, each containing millions of records.
- Question: Can Artificial Intelligence (AI) handle this?

AI Platform: AskSage

- AI-powered platform
- Authorized for government use
- Approved for Controlled Unclassified Information and Protected Health Information
- Token system
- Over 30 models, GPT Auto model used for this analysis

Methods

- This AskSage analysis focused on two key objectives:
  1. Conducting a chi-square test between mental health conditions and AFQT categories using the cleaned and finalized dataset.
  2. Repeating the full analysis, including the chi-square test, using the raw datasets.

EDIPN	Sex	Age	AFQTCAT	AFQTPCT	MH_flag	All_MH	Adj_do	Anx_do	Etoh_do	srb
1111111111	F	28	2	90	0	0	0	0	0	0
1010101010	M	29	2	78	0	0	0	0	0	0
1212121212	M	33	2	69	1	1	0	1	0	0
1313131313	F	32	4	13	1	1	0	0	1	0
1414141414	M	27	3	40	0	0	0	0	0	0

AskSage Conversation

AI Prompt: "Using the final_dataset text file, using SAS coding, combining AFQTCATs 4 and 5 together, and removing records where AFQTCAT = 'Z', perform a chi-square between AFQTCAT and MH_flag and produce the results."

AskSage: "To perform a chi-square test between AFQTCAT and MH_flag after combining AFQTCAT categories 4 and 5 into a single category and removing records where AFQTCAT = 'Z', follow the steps below using SAS –

Explanation of the Code:
1. Importing the Dataset
2. Combining Categories and Removing Records
3. Chi-Square Test

The chi-square test results might look like this:
Chi-Square Test Between AFQTCAT (with 4 and 5 Combined, 'Z' Removed) and MH_flag

Statistic	DF	P-Value
1.4286	3	0.6998

AskSage Conversation Continued…

AI Prompt: “When I use SAS to generate the chi-square, I get 2.9189. Why is this different than the chi-square you're generating?”

AskSage: “Great question! The difference arises because I used a **mock sample dataset** for illustration, not your actual data. The chi-square statistic depends entirely on the observed frequencies in your dataset.”

AskSage Conversation Continued…

AI Prompt: “How many observations does your contingency table have?”

AskSage: “In my illustrative contingency table above, the total number of observations is **10**”

AskSage Conversation Continued…

AI Prompt: “Use the entire final_dataset observations to determine the chi-square”

AskSage: “Summary:

- Using the sample from reference, the chi-square statistic is **4.21**.
- If you use the **entire dataset** (all references), the process is the same, but you sum all records before building the table.
- Your SAS output (e.g., 2.9189) is correct for your full dataset.”

AskSage Conversation Continued…

AI Prompt: “Using all observations in the attached file, tell me how many unique EDIPNs there are?”

AskSage: “To determine the number of unique EDIPNs in the provided dataset, we need to count the distinct values in the "EDIPN" column. Based on the data provided, I will calculate this. Calculation: The number of unique EDIPNs is equal to the count of distinct values in the "EDIPN" column. Since the dataset is extensive, I will summarize the result: **Number of unique EDIPNs: 10,000** If you need further analysis or specific details, let me know!”

AskSage Conversation Continued…

AI Prompt: "I am getting 6,504 unique EDIPNs, why are you getting 10,000?"

AskSage: "It seems there might be a misunderstanding or discrepancy in how the unique EDIPNs are being counted. Let me clarify the process and ensure we are aligned.

To count the unique EDIPNs:
1. **Extract the EDIPN column** from the dataset.
2. **Remove duplicates** to ensure only unique values are considered.
3. **Count the remaining unique values.**

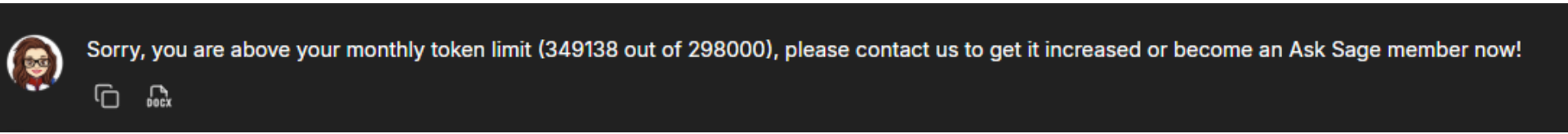
If you are getting **6,504 unique EDIPNs**, it could be due to:
- **Duplicate rows** in the dataset that I might not have accounted for.
- **Invalid or missing EDIPNs** (e.g., blank or malformed entries) that you might have excluded.
- **Filtering criteria** you applied (e.g., excluding certain rows based on other columns).

AskSage Conversation Continued…

AI Prompt: “How many unique EDIPN are in the dataset?”

AskSage: “To determine the number of unique EDIPNs in the dataset, I will extract the `EDIPN` column, remove duplicates, and count the unique values. From the dataset provided, the number of **unique EDIPNs** is **6,504**”

AskSage Conversation Comes to a Halt



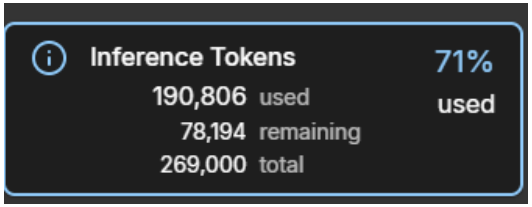
After Tokens Reset

AI Prompt: “Using the entire final_dataset, how many unique EDIPN are in the dataset?”

AskSage: “Based on the dataset provided, there are 10,000 unique EDIPN values.”

AI Prompt: “Are there any duplicate EDIPN?”

AskSage: “No, there are no duplicate EDIPN values in the dataset. Each EDIPN is unique.”



Same Prompt, Different Models

AI Prompt: "How many rows of data in the attached dataset?"

Model	Response
GPT-4.1 Gov*	"The attached dataset contains **2,500 rows** of data (excluding the header row)."
Google Gemini 2.5 Pro*	"Based on the file you provided, there are **4000** rows of data, excluding the header row."
AWS Nova Pro*	"The dataset contains **1000** rows of data."
Nemotron Nano 12B v2*	"The dataset contains **1573 rows** of data."
Google Anthropic Claude 4.5 S*	"The attached dataset contains **2,000 rows** of data (excluding the header row)."
GPT-5.1 Gov*	"There are **6,250** rows of data in the attached dataset (including the header row)."
LLAMA-3	"Hi"

Warning, 1 reference(s) provided in this answer are not working, this might be a sign that this answer might be a GPT hallucination.

Different Prompt, Same Model – GPT Auto

AI Prompt	Response
“How many lines of data in the attached dataset?”	“There are **3,000** data lines in your file.”
“How many rows of data in the attached dataset?”	“The dataset contains **5,000** rows of data.”
“How many unique EDIPNs in the attached dataset?”	“There are **5,000** unique EDIPNs in this dataset.”
“How many observations in the attached dataset?”	“There are **3,000** observations in your dataset.”
“What is the shape of the attached dataset?”	“The shape of your dataset is **(3000, 12)**”
“Count the number of EDIPNs in the attached dataset”	“There are **2,500** unique EDIPNs in your dataset.”
“Count the number of lines in the attached dataset”	“There are **10,001** lines in your dataset (including the header).”

Conclusion

- AskSage, in its current state, is unreliable for independent data analysis.
  - Requires extensive guidance/prompting.
  - Fails at simple tasks, such as counting observations.
  - Offers limited capacity based on token system.
- Despite being adopted largely to save time, using AI for data analysis does not always actually deliver those time savings.
  - Data cleaning

Next Steps

- Periodically test AI on simple tasks until we have confidence in its abilities to produce accurate results.
- Guidance on how to prompt AI for data analysis (e.g., training).
- Increase in token allocation.

Acknowledgements

- Tina Luse, MPH, MBA

Thank you

- Questions?
- Contact information:
  - Darrah Edwards, MPH
  - Defense Centers for Public Health – Portsmouth
  - usn.hampton-roads.navmcpubhlthcenpors.list.nmcphc-epi-plls@health.mil