AD-AutoGPT: An Autonomous GPT for Alzheimer's Disease Infodemiology

Authors

Haixing Dai[1], Yiwei Li[1], Zhengliang Liu[1], Lin Zhao[1], Zihao Wu[1], Suhang Song[2], Ye Shen[3], Dajiang Zhu[4], Xiang Li[56], Sheng Li[7], Xiaobai Yao[8], Lu Shi[9], Quanzheng Li[56], Zhuo Chen[2], Donglan Zhang[10], Gengchen Mai[81], Tianming Liu[1]

Review By-:

Dev Rishi Verma

For fulfilment of internship assignment for

CogAI4Sci Lab NUS

24.10.2024

Paper Summary & Review

Overview: This paper presents **AD-AutoGPT**, a tool inspired by **AutoGPT**, an open-source application based on the **GPT-4 mode**l. Users can input text prompts, and the system autonomously gathers, processes, analyses and saves intricate health data regarding Alzheimer's Disease.

Problem to be addressed: Alzheimer's Disease presents a critical challenge, requiring heightened public awareness and data-driven insights to address its complexity. Public health professionals often depend on labour-intensive techniques like web scraping, API data collection, and manual data processing to extract meaningful information from news media and health reports. With the increasing global data volume, these efforts become more difficult, making it crucial to develop streamlined, automated solutions.

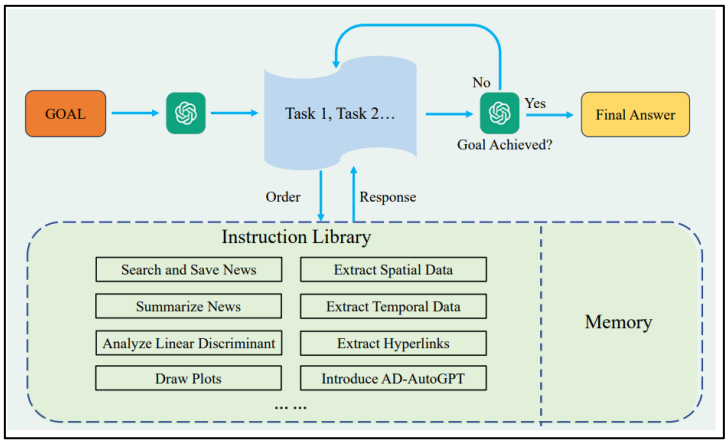
AutoGPT, while promising, attempts to simplify some of these tasks by automating data collection and analysis. However, despite its potential, the model faces issues such as inconsistent handling of unstructured data, limited contextual understanding for nuanced topics, and a reliance on substantial computing power, which can limit its widespread application. This highlights the need for continued improvements to ensure it meets the diverse needs of the public health community.

Aspect	AutoGPT Limitations	AD-AutoGpt Improvements
Information retrieval	Limited ability to acquire specialized information quickly and precisely, especially for Alzheimer's Disease (AD).	Incorporates specific prompting mechanisms to enhance efficiency and relevance of information acquisition from authoritative sources.
Detail Extraction	Challenges in accurately extracting critical details like timestamps and geolocation from news articles.	Uses web-crawling scripts and geo-location libraries for precise detail extraction.
Token Limitations	Restricted by a token limit (e.g., 4096 tokens), leading to missing essential content and details.	Segments and vectorizes text for independent processing, allowing comprehensive summaries beyond the token limit.
Summary Analysis	Lacks capacity for in-depth analysis of generated summaries, resulting in potentially redundant or incomplete information.	Employs Latent Dirichlet Allocation (LDA) to extract pertinent keywords, providing clarity on central themes in the AD domain.
Visualization Capabilities	Does not offer robust visualization options for effectively representing data.	Integrates effective visualization techniques for creating plots representing news occurrences, geographical highlights, and the evolution of research keywords over time.

A comparison table illustrating how AD-AutoGPT outperforms general AutoGPT in addressing Alzheimer's Disease-specific user queries.

AD-AutoGpt Architecture:

In this section I'll summarize the architecture of AD-AutoGpt



Framework of AD-AutoGPT as given in the paper

AD-AutoGPT features a specialized **instruction library** that includes tailored functions and tools for public health infodemiology tasks, and it can be expanded and updated as needed.

Some of these are:

1. **Search and Save News:** Utilizes **Google API** to search for the latest news from authoritative websites and saves the URLs on a local device.
2. **Summarize News:** Uses ChatGPT or GPT-4 to summarize the main content of news articles and extract spatial-temporal information from each stored news piece.
3. **Visualize Results:** Draws visual representations of the results and displays **LDA analysis** outcomes of the news content.

The cognitive process of AD-AutoGPT can be illustrated through a structured set of prompts. Each task is guided by an input prompt that comprises these essential components:

Components of Input Prompt:

1. **Question**: The problem the AI needs to solve.
2. **Thought**: The reasoning and thought process the AI uses to tackle the problem.
3. **Action**: The operation the AI selects, which it believes is the best solution to the task.
4. **Action Input**: The input provided to the function for execution.

Components of Output Prompt:

1. **Observation**: The result of the function, used to guide the AI's next step.
2. **Thought**: The AI's evaluation of the observation and how it influences its next move.
3. **Final Answer**: The conclusive judgment, with an iterative loop continuing until a satisfactory solution is achieved.

Functionality

AD-AutoGPT includes crucial functions for optimising data analysis and insight production in Alzheimer's disease research. Here's a brief introduction of its main capabilities:

1. **Efficient Summarization** AD-AutoGPT extracts key information from news articles using GPT-4 for text summarization. It crawls news URLs, retrieves the content, and summarizes it using a map-reduce approach, ensuring efficient processing of large texts.
2. **Spatiotemporal Information Extraction** The model extracts temporal data from news metadata and spatial data through the **GeoText** geoparsing tool, linking place names to geographic coordinates.
3. **LDA Analysis** Latent Dirichlet Allocation (LDA) is applied for topic modelling, clustering documents based on frequent keywords, and offering insights into main article themes.

Case Study and Experimental Results

Alzheimer's Disease News Information Retrieval:

AD-AutoGPT demonstrated its capabilities by autonomously gathering **277 news articles** from reputable Alzheimer's sources over the course of a year. The system effectively handled all aspects of the process, including text extraction, summarization, spatiotemporal analysis, topic modelling, and visualization, without requiring any human oversight.

Visualization of Spatiotemporal Data:

The model adeptly visualized the geographical distribution of Alzheimer's-related articles, highlighting a **significant concentration in the U.S. and Western Europe.** Furthermore, a temporal analysis revealed a notable surge in coverage during September 2022, illustrating the model's ability to track trends over time.

LDA Topic Modelling and Hot Topic Analysis:

Utilizing Latent Dirichlet Allocation (LDA), AD-AutoGPT successfully identified critical research topics and monitored keyword trends. This functionality allows users to quickly comprehend prevailing research trends without the need to sift through lengthy reports, streamlining the process of extracting valuable insights from vast amounts of information.

Conclusion

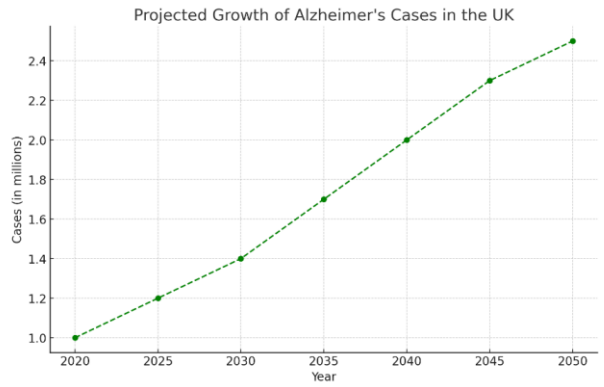
In conclusion, AD-AutoGPT revolutionizes Alzheimer's research by streamlining the processes of data collection, analysis, and visualization, ultimately reducing the burden on researchers. By automating these tasks, the system not only enhances efficiency but also ensures that critical insights are prioritized over the tediousness of manual data handling. This innovation paves the way for more effective public health interventions and facilitates a deeper understanding of Alzheimer's Disease dynamics. With AD-AutoGPT, researchers can devote their expertise to interpreting data and implementing findings, thereby accelerating the development of solutions and improving outcomes for those affected by this complex condition.

Potential Improvements

1. In my view, while the visualizations offer useful insights for exploratory data analysis, they fall short in delivering precise information that directly addresses the user's query. There's significant room for improvement. Instead of relying on predefined commands to generate static plots, we could leverage GPT-4 itself to create more dynamic, query-driven visualizations. This would enhance user experience by tailoring the visuals to answer specific questions, providing clearer, more actionable insights rather than just generalized data representations.

   To test this, I fed my article summaries directly into ChatGPT since I couldn't access the API due to its paid nature:

   Outputs:

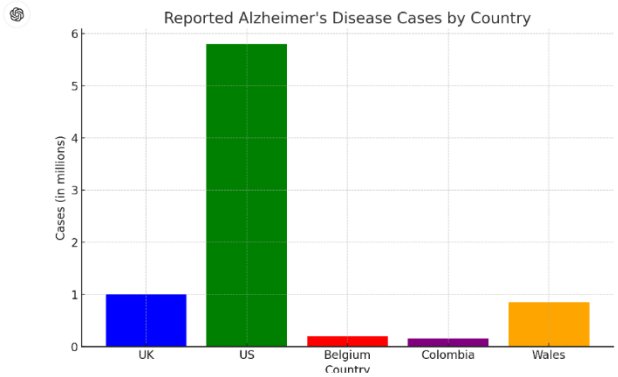


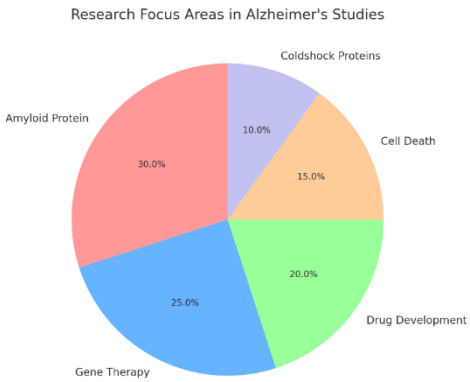
the NHS)
Summary:
Sally Magnusson, a journalist whose mother suffered from Alzheimer's disease, grapples with the possibility of inheriting the condition. She explores the implications of a new blood test that can detect the presence of amyloid, a protein associated with Alzheimer's. The test offers hope for early detection and potentially preventing the disease before symptoms develop, but raises concerns about the psychological impact of knowing one's future. Magnusson's film "Alzheimers, a Cure and Me" explores these complexities, highlighting the need for more research participants. The film emphasizes that while scientists are optimistic about potential cures, more people need to participate in research trials to develop effective treatments. The film will be broadcast on BBC Scotland at 2100 on Sunday, 13 October, and will also be available on the iPlayer.

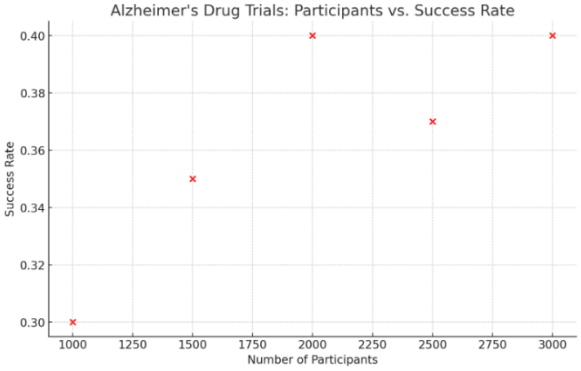
**

using this information draw some intresting plots and graphs

3/3

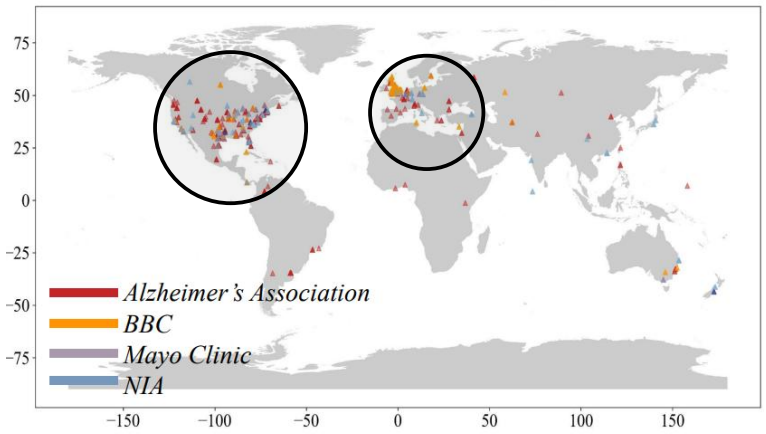






These plots can be enhanced with more refined prompts, which I believe will allow users to grasp the context of the summaries more effectively. By tailoring the visuals to directly address the user's specific query, the insights provided will become clearer and more actionable.

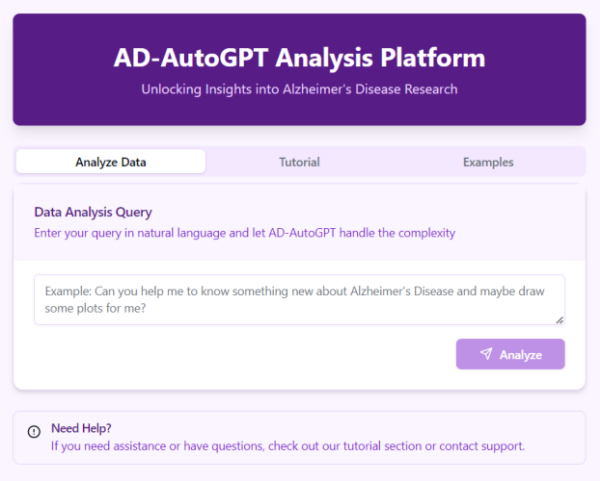
2. AD-AutoGPT currently relies on a limited set of news sources, including BBC, Mayo Clinic, Alzheimer's Association, and the National Institute on Aging (NIA), all of which are based in the US and UK. This narrow focus restricts the global scope of the data, as evidenced by the visualization map, where Asian countries and other regions are sparsely represented, while the US and UK show a dense concentration of data points. Incorporating a more diverse range of news channels and also incorporating research articles from different regions would provide a more comprehensive and accurate global perspective on Alzheimer's disease trends and discussions.

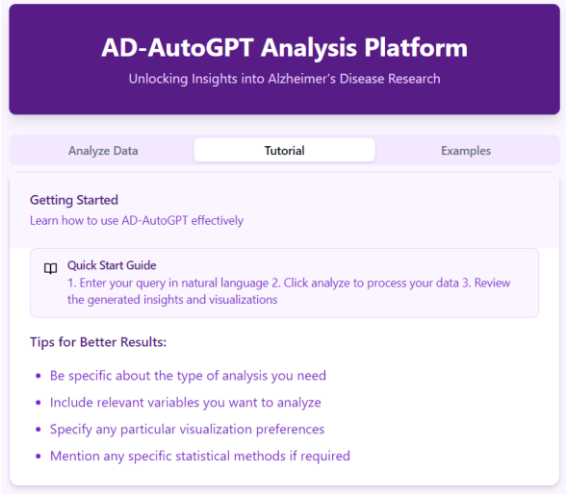


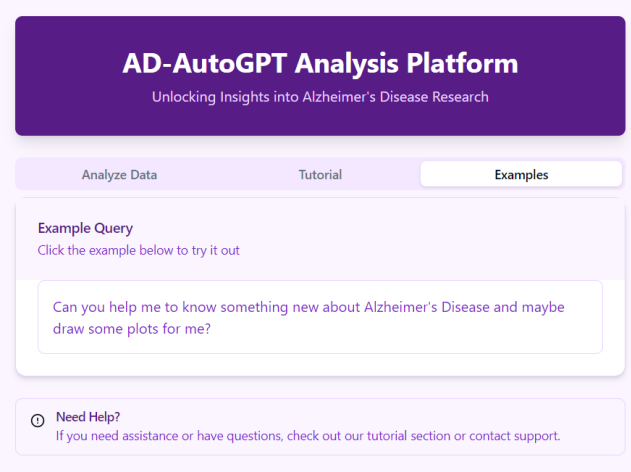
3. I believe that distinguishing between the needs of professionals and the general public can greatly improve the effectiveness of the tool. For example, health professionals and data analysts would likely benefit from more detailed statistical content and comprehensive global analyses, offering deeper insights into current developments and responses. Meanwhile, the general public may prefer simpler explanations, focusing on local news and key points, with visualizations that are clear and easy to interpret. By catering to the specific needs of each group, we can deliver more relevant and accessible information to a wider audience.

4. One key issue of using AD-AutoGPT is its complexity, which makes it difficult for the average user to navigate and apply properly. Given this, making it a user-friendly website could significantly improve accessibility. This platform would allow consumers to simply enter their queries and obtain responses, without having to comprehend the underlying technological details. A web interface might incorporate guided workflows, intuitive visualisations, and built-in tutorials, changing AD-AutoGPT from a technical tool to a useful resource for public health practitioners and researchers. This technique would allow a larger audience to profit from the benefits of advanced data analysis without having to deal with the complexities of the underlying model.

I personally designed some user-friendly web interfaces for AD AUTO-GPT. These interfaces make it incredibly easy to access and use the application, allowing us to reach a much broader audience.







PUBLIC SURVEY

To analyse and understand the research question and problem more effectively, I conducted a survey involving 53 participants. Their data was collected and analysed to gain insights. It's important to note that the participants in this survey come from similar backgrounds and demographics, with the majority residing in Jaipur and belonging to Indian families. This shared context provides a focused perspective for the analysis.

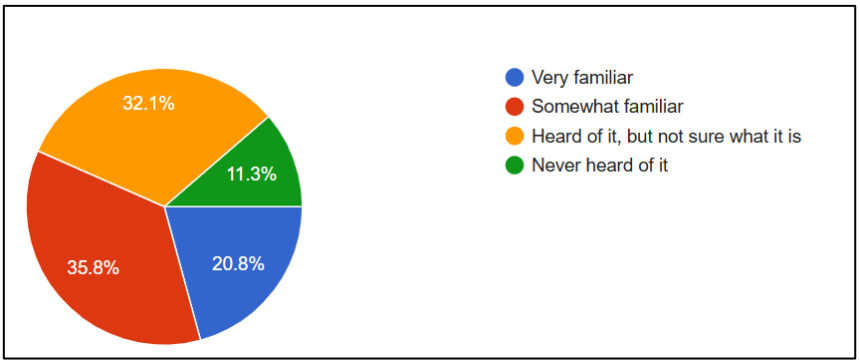
I have discussed some of the important insights of the survey below which helps us to understand the need of AD AUTO-GPT better

The link to the complete excel sheet of the survey can be found here:

https://docs.google.com/spreadsheets/d/1swvO4x5z9o9E54YtC4U4w-6DYy-WW10UARBkqyHESG8/edit?usp=sharing

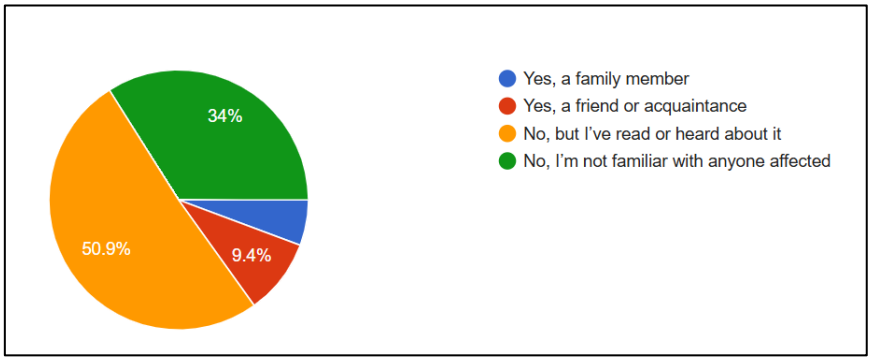
Question:

How familiar are you with Alzheimer's Disease?



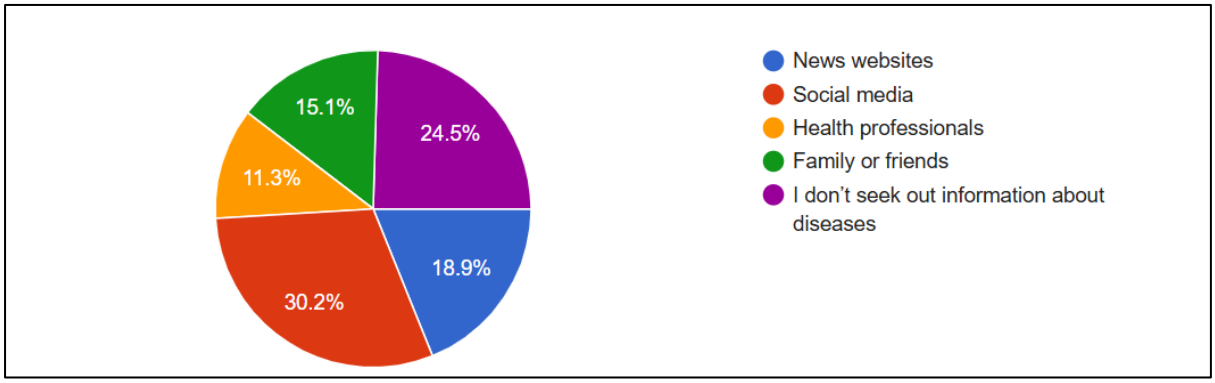
Despite the extreme danger posed by Alzheimer's Disease, only 20.8% of people are very familiar with it. The majority are either somewhat familiar or have merely heard of it, while 11.3% have never heard of it at all.

Do you know anyone personally who has been affected by Alzheimer's Disease?



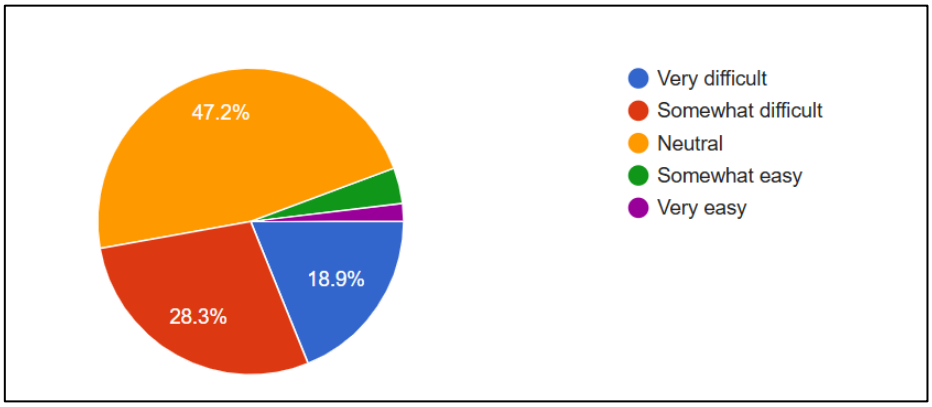
there is a degree of awareness regarding Alzheimer's Disease, many individuals lack direct personal connections to it. By addressing this familiarity gap through education and community engagement, we can foster a more informed and compassionate understanding of the disease and its profound effects on individuals and families.

How do you usually get information about diseases like Alzheimer's?



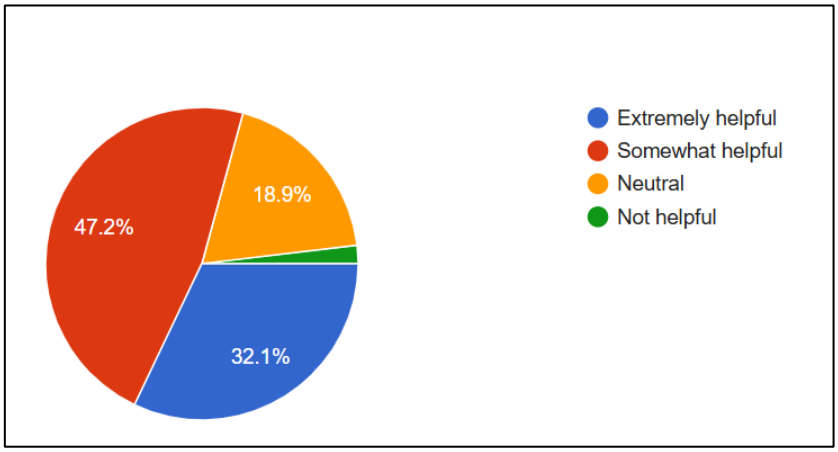
Social media being the primary source (30.2%) for health information is problematic as it can spread unverified medical claims and misinformation rapidly, while lacking the credibility and accuracy of professional medical guidance.

When reading about health issues like Alzheimer's Disease, how difficult is it to understand complex data or medical reports?



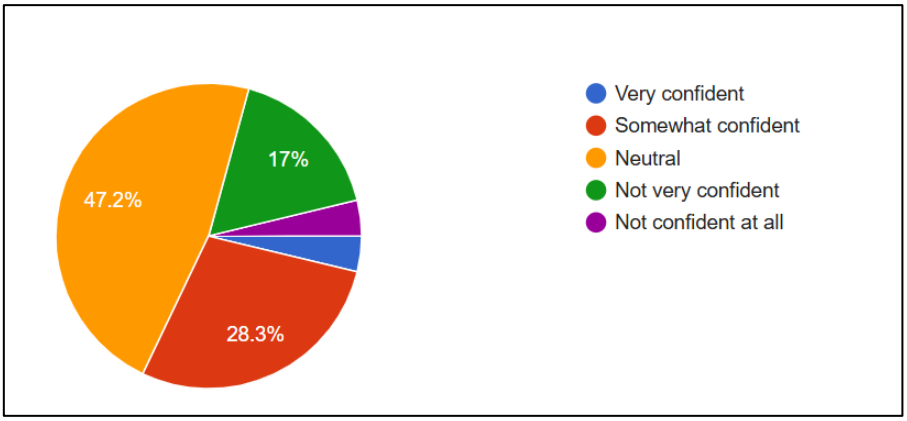
This indicates a clear problem: nearly half of people struggle with understanding complex medical information about Alzheimer's, while the other half are neutral, suggesting medical reports and data aren't effectively reaching the general public. The tiny percentage finding it easy (5%) highlights a significant gap between expert communication and public comprehension, pointing to a need for more accessible, plain-language medical communication.

If there was an AI tool that could automatically gather and analyze global Alzheimer's research and news from trustful sources, how helpful would that be for you and society?



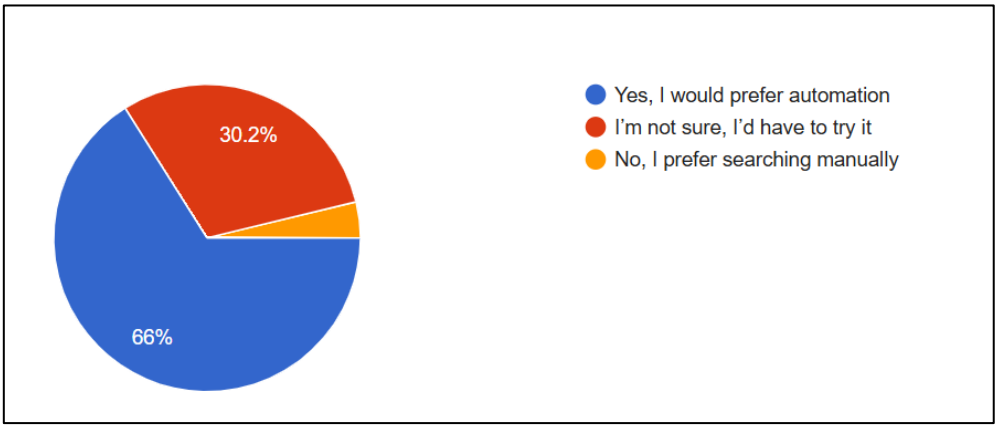
The data reveals strong public support for an AI tool to analyze Alzheimer's research, with nearly 80% viewing it positively (47.2% somewhat helpful, 32.1% extremely helpful), while only 18.9% remain neutral and a minimal ~2% find it unhelpful - suggesting widespread recognition of AI's potential to bridge the gap between complex medical research and public understanding, especially significant given earlier data showing many struggle with comprehending medical information.

How confident are you in current methods of gathering health information (such as news sites, social media, etc.) to provide accurate, global perspectives on diseases like Alzheimer's?



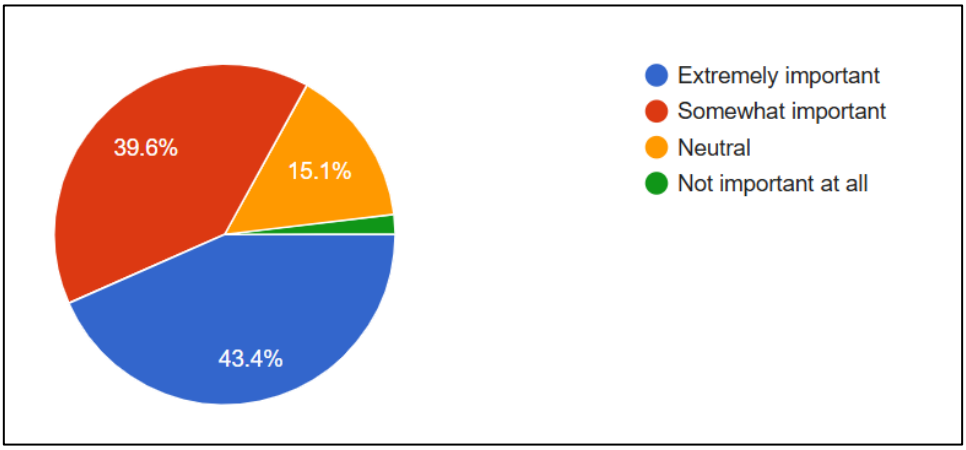
very few respondents lack confidence in current methods, with most either expressing strong confidence or remaining neutral. This highlights an urgent need for software that provides quick and reliable healthcare insights.

Would you prefer a platform that automates the process of gathering and summarizing health-related data instead of manually searching for news and research articles?



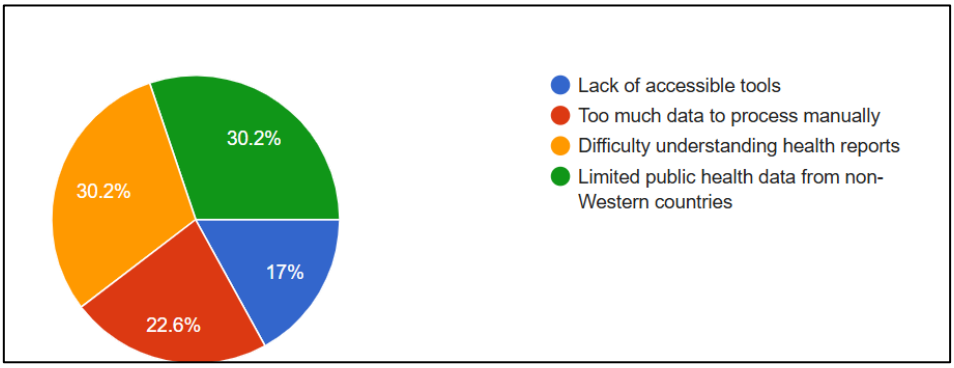
The survey shows a strong preference for automation in gathering and summarizing health-related data, with 66% of respondents favouring an automated platform like AD AUTO-GPT. 30% are unsure and would need to try it first, while very few prefer manual searches. These results highlight a clear demand for tools that enhance efficiency in accessing health information, suggesting a significant opportunity for automation in this field.

How important is it for you to access health insights from a variety of countries and regions, not just the US or UK?



Approximately 90% of respondents believe it's essential to access health information from diverse global news sources instead of relying solely on outlets from the UK and US. This highlights the need for the AD AUTO-GPT platform to broaden its scope in gathering information.

In your opinion, what are the biggest barriers preventing people from understanding global trends in diseases like Alzheimer's?



The equal distribution of responses highlights significant issues: the lack of available tools, the overwhelming volume of data to process manually, difficulties in understanding health reports, and limited public health data from non-Western countries. These challenges need to be addressed to enhance people's understanding of global disease trends, such as Alzheimer's. This underscores the urgent need for AI-enhanced tools like AD: AUTO-GPT.

In conclusion, the survey results underscore a critical need for improved awareness and understanding of Alzheimer's Disease among the public. Despite the severity of the issue, only a small percentage of individuals are very familiar with it, highlighting a significant knowledge gap. While social media serves as a primary information source, it often spreads misinformation, complicating public comprehension of complex medical data. The strong support for AI tools like AD AUTO-GPT reflects a collective recognition of the potential for technology to bridge the divide between intricate research and accessible health information.

Furthermore, the overwhelming demand for diverse, global health insights emphasizes the necessity for platforms that can aggregate and summarize information from a variety of credible sources. Addressing barriers such as the lack of tools and the challenge of deciphering health reports is essential for fostering a more informed and compassionate society. By investing in AI-enhanced solutions like AD AUTO GPT itself and prioritizing education and engagement, we can empower individuals and communities to better understand the global trends in diseases like Alzheimer's, ultimately improving the quality of care and support for those affected.