

From Diaries to Digital: The Role of AI in Web-Mediated Documentary Analysis

Laura Arosio

University of Milano-Bicocca, Milan, Italy

This paper explores how artificial intelligence (AI) can support social researchers in utilizing web-mediated documents for research purposes. It extends traditional documentary analysis to include digital artifacts such as blogs, forums, emails and online archives. The discussion highlights the role of AI in different stages of the research process, including question generation, sample and design definition, ethical considerations, data analysis, and results dissemination, emphasizing how AI can automate complex tasks and enhance research design. The paper also reports on practical experiences using AI tools, specifically ChatGPT-4, in conducting web-mediated documentary analysis and shares some ideas for the integration of AI in social research.

Keywords: artificial intelligence, generative AI, web-mediated documents, documentary analysis, data analysis with AI, social research methodology

Introduction

The digital age is transforming the landscape of social research, impacting both what is studied and how it is studied. On one hand, new communication technologies have created new spaces where individuals spend their time, perform tasks, interact with each other, and share meanings. On the other hand, these innovations provide strategies for conducting research, particularly through the integration of advanced technologies such as artificial intelligence (AI).

Documentary analysis is a classical method in social research that involves extracting meaningful information from various types of materials not originally intended for research purposes, such as diaries, letters, laws, films, books, songs, and regulations. The advent of the internet has expanded this domain to include web-mediated documents (Arosio, 2010), such as blogs, emails, personal and institutional websites, forums, and online archives. These digital artifacts provide a rich source of data, reflecting ideas, values, and social perspectives. However, the volume and complexity of online content pose significant challenges for researchers.

This paper aims to explore how AI can support social researchers in utilizing web-mediated documents for research purposes. AI can assist not only in the data analysis, but throughout the entire research process, from the initial conceptualization of the study to the dissemination of results. By using AI techniques, researchers can improve the research design and automate labor-intensive processes, facing challenges hidden in vast digital archives. This integration not only offers continuity with traditional documentary analysis but also introduces unprecedented opportunities in the field of social research.

Laura Arosio, Ph.D., Assistant Professor in Sociology, Department of Sociology and Social Research, University of Milano-Bicocca, Milan, Italy.

The paper begins by highlighting the contribution of documentary analysis to social research and introduces web-mediated documents (Section 2). The discussion then moves to the potential contribution of artificial intelligence (AI) in the different phases that researchers go through when conducting research with computer-mediated documents (Section 3). Section 4 focuses on data analysis, highlighting how documents can be used in both qualitative and quantitative research approaches. In Section 5, a report is presented on a working session conducted with an AI program (specifically, ChatGPT-4) to perform web-mediated documentary analysis. Finally, Section 6 shares impressions that emerged from the use of AI tools for conducting web-mediated documentary research.

Traditional Documents and Web-Mediated Documents

In social research, the analysis of documents has long been a foundational method, providing rich insights into human behavior, societal norms, and cultural practices (Scott, 1990; Prior, 2003). The contribution of documentary analysis was highly valued by the founders of sociology, including Emile Durkheim, Max Weber, and Karl Marx (see McCulloch, 2004).

“Traditional” documents encompass a wide variety of materials. Personal documents, such as diaries, letters, autobiographies, and family photographs, offer intimate insights into individual lives and personal experiences (Plummer, 1983). Institutional documents, including regulations, laws, meeting minutes, and official reports, reveal the inner workings and formal communications of organizations. Media documents, like press articles and radio and television programs, capture the dissemination of information and cultural narratives over time. Cultural documents, ranging from literary works and artistic creations to folklore, preserve the artistic and symbolic heritage of societies.

Documents have been crucial in social research, providing contextualized data that help researchers understand historical events, cultural shifts, and personal journeys. With the rise of the internet, a new category of documents has emerged: web-mediated documents (Arosio, 2010; 2022). These digital artifacts, including blogs, emails, personal and institutional websites, forums, and online archives, create new dimensions of interpretation for researchers.

Web-mediated documents offer several advantages for social research. They provide dynamic and diverse data sources that capture current trends, public opinion, and evolving cultural narratives. For example, blogs and social media platforms allow individuals to share their thoughts and experiences instantly. Institutional websites and online reports offer accessible information on organizational practices and policies, while digital media platforms provide a continuous stream of news and cultural content.

However, the volume and complexity of online material pose significant challenges for researchers. The vast amount of data available on the internet require sophisticated tools and methods for collecting data and effective analysis. Issues of authenticity, such as verifying the credibility of sources and distinguishing genuine content from misinformation, are critical concerns. Additionally, ethical considerations regarding privacy and consent are amplified in the digital realm, where personal information is often publicly shared without explicit permissions. Another challenge concerns the selection of the documentary corpus (the issue of sampling), in both a qualitative and quantitative perspective.

The Role of Artificial Intelligence in Web-Mediated Documentary Analysis

Artificial intelligence (AI) promises to transform the way researchers conduct social research (Grossmann

et al., 2023). In a recent work, Xu and colleagues (2024) summarized literature dealing with the application of AI at every stage of social science research. AI can serve as an effective tool for social science research, including tasks such as literature searching and reviewing (McGee, 2023), generating questions and hypotheses (Banker, Chatterjee, Mishra, & Mishra, 2023; Park et al., 2024), analyzing data (Ziems et al., 2023), and assisting with writing (Chen, 2023).

As far as web-mediated documentary analysis is concerned, it is plausible that AI could impact different stages of the research process, such as defining research questions and theoretical context, selecting documents and sampling, designing the research, analyzing documents, interpreting results, disseminating findings, and addressing ethical issues.

Defining Research Questions and Theoretical Context. AI can assist researchers in formulating research questions by analyzing existing literature and identifying gaps in knowledge. For example, an AI system could analyze thousands of academic papers on a specific topic, identifying under-researched areas and suggesting new research questions. By proposing alternative ideas, AI helps researchers choose the questions to explore. Text mining and bibliometric analysis tools can also play a role in contextualizing new research within the current academic discourse.

Selecting Documents and Sampling. One of the critical challenges in web-mediated documentary analysis is selecting a sample from a vast pool of potential sources. AI-driven tools can suggest effective sampling strategies and automate the process of document retrieval. For instance, AI can help identify and collect documents from online databases, forums, and websites through techniques such as web scraping and automated content filtering. Machine learning algorithms can then ensure the quality of the process, whether the study requires probabilistic or non-probabilistic samples.

Defining the Research Design. AI can assist in making research design choices, such as determining the temporal scope of the study, crafting questionnaire items, and developing critical interrogation guides for web-mediated documents. For example, an AI system could analyze trends in social media discussions over the past decade to help researchers decide on the most relevant time frame for their study.

Analyzing Documents. AI promises to enhance the analytical capabilities of researchers. AI can process and analyze large volumes of data much faster and more precisely than manual methods. For example, AI can analyze thousands of online reviews to identify common themes and sentiments about a product or service. AI can identify and classify entities (e.g., names, dates, locations) within online texts, uncover hidden themes, and condense long texts into concise summaries. AI promises to perform and execute initial quantitative data analyses.

Interpreting Results. AI tools can help interpret results by providing statistical and thematic analyses. Visualization tools further aid in presenting complex data in an understandable format, making it easier for researchers to draw meaningful conclusions.

Dissemination of Findings. AI can assist in the dissemination of research findings by generating automated reports tailored for different audiences, translating academic findings into more accessible language for policymakers and the general public. For example, AI can produce a summary of a complex research study on housing policies conducted through citizen feedback on an online platform, highlighting to policymakers the most critical findings and recommendations.

Ethical Considerations. The use of web-mediated documentary analysis raises important ethical issues. AI can help ensure that ethical standards are met. For instance, AI can analyze the language used in online media

reports to detect potential biases and ensure a balanced representation of different viewpoints. It can also assist in maintaining the privacy and confidentiality of sensitive information contained in web-mediated documents.

AI and the Analysis of Web-Mediated Documents

One of the most discussed capabilities of artificial intelligence in social research is its potential to perform data analysis, which in web-mediated documentary research can be either qualitative or quantitative (depending on the research questions and theoretical framework).

AI can automate the collection of large datasets from the web using web scraping and data mining techniques. Moreover, AI tools can clean and pre-process the data by removing duplicates, filtering out irrelevant content, and standardizing formats.

AI can be employed to identify and categorize entities such as people, organizations, and locations within the documents, detect the emotional tone of the online texts, and reveal underlying themes and trends in web content. Machine learning algorithms can classify and categorize web-mediated documents based on predefined criteria. For example, AI can group social media posts by topic, sentiment, or source credibility.

AI can also map and analyze the relationships and interactions between individuals in online communities. By examining patterns of communication and influence within networks, researchers can identify key influencers, study the spread of information, and understand the dynamics of online communities.

AI can support the process of quantitative data analysis on documents mediated by the internet. For instance, AI can be employed to collect and extract data from online sources such as websites, forums, and social platforms. Once the data are extracted, AI can perform descriptive statistical analyses, and even provide the necessary syntax for more advanced analyses using tools like R or Python. Additionally, AI can create visual representations of the data, including graphs, charts, and heat maps. Moreover, AI can assist in building predictive models and simulations based on historical data, offering insights into potential future trends.

In general, it is important to emphasize that, at present, AI does not autonomously perform complex analyses neither qualitative nor quantitative. Instead, AI provides support by helping to design analysis frameworks, offering step-by-step instructions, and suggesting the most suitable programs to use. It also writes the syntax commands, which is often one of the most technically challenging task for social researchers. The scenario is constantly evolving, and it is not out of the question that we may soon see the direct integration of complex analysis tools into AI.

My Experience Using AI for Web-Based Documentary Analysis

After a theoretical exploration of the potential of artificial intelligence in documentary analysis, I decided to test these tools. The goal was to understand not only the advantages but also the practical challenges that arise when integrating AI into web-based documentary analysis processes. In the following paragraphs, I provide an account of the working session where I asked ChatGPT-4 to assist me in various stages of my research with web-mediated documents.

The premise is the crucial importance of designing effective prompts. A prompt is a specific input or question given to the AI, designed to elicit a particular type of response. The quality of the prompt significantly impacts the AI's output. Effective prompts should be specific, avoid ambiguity, and provide enough context to guide the AI's understanding. Characteristics of well-constructed prompts include clarity, brevity, and relevance to the desired outcome (OpenAI, 2024).

One effective strategy is to specify the role ChatGPT should assume and clearly state the objectives of the task. For example, a good prompt for analyzing web-mediated documents might be: “As a social researcher, analyze the recurring themes of community identity and political engagement in online forums and blogs discussing climate change”. This prompt is specific, provides clear context, directs the AI towards a focused analysis, and sets a clear objective. In contrast, a poor prompt would be: “Tell me about online forums”. This vague prompt lacks specificity and context, leading to broad and potentially irrelevant responses.

I will share the prompts I used in my work session and the responses ChatGPT-4 provided. In some cases, for the sake of brevity and readability, I will include only portions of the responses and offer a summary.

As a first request, I asked ChatGPT to generate ideas for a research study using web-mediated documentary analysis. The session resulted in five research study proposals, each demonstrating diversity in terms of the themes addressed, the types of documents analyzed, and the research methods employed (Table 1).

Table 1

ChatGPT Generates Ideas for a Research Study Using Web-Mediated Documentary Analysis

<p>Prompt: Imagine you are a social scientist wanting to conduct documentary analysis on web-mediated documents. Generate ideas for a research study.</p> <p>Response: Research Study 1: Analysis of Social Media Discourse on Environmental Sustainability Objective: To explore how social media users discuss environmental sustainability and identify the main themes, concerns, and sentiments expressed. Research Questions: What are the recurring themes in social media posts regarding environmental sustainability? What sentiments (positive, negative, neutral) dominate the discussions on sustainability? How do discussions and sentiments vary by geographic region or demographic group? Methodology: Data Collection: Use web scraping to collect posts from Twitter, Facebook, and Instagram with relevant hashtags such as #Sustainability, #Environment, #EcoFriendly. Data Pre-processing: Clean the data by removing spam, duplicates, and irrelevant posts. Analysis: Use NLP for sentiment analysis and topic modeling. Map the discussions and identify the main named entities. Interpretation: Contextualize the results within the theoretical framework of environmental sustainability. Dissemination: Create a report with graphical visualizations of the predominant themes and sentiments.</p> <hr/> <p>Research Study 2: Exploration of Narratives in Online Forums about Mental Health Objective: To analyze how users of online forums discuss and share experiences about mental health. Research Questions: What are the main concerns and challenges expressed by users regarding mental health? What resources and coping strategies are mentioned and shared? How do discussions evolve over time? Methodology: Data Collection: Collect data from specific online forums about mental health (e.g., Reddit, HealthBoards). Data Pre-processing: Anonymize the data and remove non-relevant content. Analysis: Use text mining techniques to identify the main themes. Analyze the frequency and content of the discussions. Interpretation: Place the narratives within the context of mental health theories and social support. Dissemination: Publish a report and present the findings at academic conferences.</p>

Research Study 3: Monitoring Public Opinion on Immigration Policies through Social Media Platforms

Objective: To monitor and analyze public opinions on immigration policies through social media platforms like Twitter, Facebook, and Instagram.

Research Questions:

What are the main arguments for and against immigration policies?

What sentiments (positive, negative, neutral) dominate the discussions on immigration policies?

How do opinions vary by geographic region or demographic group?

What specific events or political changes most influence the discussions?

Methodology:

Data Collection: Use web scraping techniques to collect posts from Twitter, Facebook, and Instagram with relevant hashtags such as #ImmigrationPolicy, #ImmigrationReform, and #ImmigrantRights.

Data Pre-processing: Clean the data by removing spam, duplicates, and irrelevant posts. Normalize the text to remove linguistic variations and standardize the format.

Analysis:

Use Named Entity Recognition (NER) to identify and categorize relevant entities such as politicians, organizations, and events.

Conduct sentiment analysis to assess the emotional tone of the posts to determine whether sentiments are positive, negative, or neutral.

Apply topic modeling algorithms to uncover the main themes and arguments discussed in the posts.

Perform temporal analysis to analyze how opinions evolve over time in relation to specific political events or changes.

Interpretation: Contextualize the results within the theoretical framework of political communication and public opinion theories. Compare the results with previous studies to identify continuity and changes in public discourse.

Dissemination: Present the results through a detailed report and graphical visualizations (e.g., bar charts for theme distribution, heat maps for regional sentiments). Publish the results in academic journals and present at political science conferences.

Research Study 4: Analysis of Blog and Opinion Article Discussions on Climate Change

Objective: To explore how blogs and opinion articles discuss climate change, identifying main themes, concerns, opinions, and sentiments expressed.

Research Questions:

What are the recurring themes in blog and opinion article discussions about climate change?

What sentiments (positive, negative, neutral) dominate the discussions on climate change?

How do discussions and sentiments vary by source (independent blogs, traditional media, etc.)?

What specific events (e.g., climate conferences, natural disasters) most influence the discussions?

Methodology:

Data Collection: Use web scraping to collect blogs and opinion articles from news websites, independent blog platforms, and other relevant sources with keywords such as “climate change”, “global warming”, “climate action”, and “environmental policy”.

Data Pre-processing: Clean the data by removing spam, duplicates, and irrelevant content. Classify the articles by source and date.

Analysis:

Use NLP for sentiment analysis and topic modeling.

Map the discussions and identify the main named entities.

Interpretation: Contextualize the results within the theoretical framework of climate communication and public opinion theories. Compare the results with previous studies to identify continuity and changes in public discourse.

Dissemination: Create a report with graphical visualizations of the predominant themes and sentiments. Publish the results in academic journals and present at specialized climate and communication conferences.

Research Study 5: Examination of Consumer Behavior through Online Reviews on E-commerce Platforms

Objective: To analyze consumer behavior and sentiment through online reviews on e-commerce platforms like Amazon, eBay, and Yelp.

Research Questions:

What are the main factors that influence positive and negative reviews?

How do consumers express their satisfaction or dissatisfaction in online reviews?

What trends and patterns can be identified in consumer behavior over time?

Methodology:

Data Collection: Use web scraping to collect reviews from e-commerce platforms like Amazon, eBay, and Yelp for a variety of products and services.

Data Pre-processing: Clean the data by removing spam, duplicates, and irrelevant reviews. Normalize the text to remove linguistic variations and standardize the format.

Analysis:

Use sentiment analysis to assess the emotional tone of the reviews.

Apply topic modeling to uncover the main themes and factors discussed in the reviews.

Perform trend analysis to identify changes in consumer behavior over time.

Interpretation: Contextualize the results within consumer behavior theories and marketing strategies. Compare the results with previous studies to identify patterns and trends.

Dissemination: Create a report with graphical visualizations of the predominant themes, sentiments, and trends. Publish the results in marketing journals and present at consumer behavior conferences.

These five research study ideas provide a structured approach to analyzing web-mediated documents on various important social topics. Let me know if you need further modifications or additional details!

I asked ChatGPT to continue with topic 4, which focuses on analyzing discussions in blogs and opinion articles about climate change. I decided to handle the literature collection by myself, but I requested assistance in selecting the most appropriate data for conducting this research. I was advised to start with a selection of Italian sources, categorized into three main areas: Major Italian News Websites, Popular Italian Environmental Blogs, Academic Platforms and Independent Blogs. Next, I was advised to define specific keywords to refine the search and gather relevant content. Finally, I was provided with techniques for data collection, such as using Python libraries for web scraping and checking the availability of APIs on the selected platforms to facilitate data access (Table 2). When I asked ChatGPT to select documents focused on blog and opinion article discussions about climate change, it informed me that it could not do so directly. However, it guided me step-by-step through the installation of programs and provided the necessary syntax/commands.

I asked for further ideas about Case 4 on sampling, ethical considerations and data analysis (Table 3). ChatGPT responded by breaking down the approach into three key sections. First, the sampling strategy was outlined. The strategy included using stratified and random sampling techniques to ensure a wide range of perspectives from various sources. The response highlighted the importance of covering a consistent time period to capture trends and changes over time. Second, the ethical considerations were addressed, stressing the need for privacy protection, proper data storage, and respect for the original content. ChatGPT emphasized the importance of anonymizing any personal information and ensuring that the use of content adhered to fair use policies. The response also discussed the need to handle sensitive topics with care, aiming for balanced representation to avoid bias or sensationalism (Table 3).

Finally, ChatGPT described how thematic analysis and content analysis could be used to identify and interpret key themes within the articles, supported by software tools like NVivo or ATLAS.ti. On the quantitative side, the response included methods such as sentiment analysis, trend analysis, and correlation analysis, with suggestions for using natural language processing (NLP) tools to automate the analysis and visualize the results through graphs and charts (Table 3).

Table 2

ChatGPT Gives Assistance in Selecting the Most Appropriate Databases and Sources for Conducting Research on a Selected Topic

<p>Prompt: I want to continue with topic 4. I will take care of the literature collection. You help me select the database.</p> <p>Response: Sure, let's proceed with selecting the appropriate databases and sources for your research on the analysis of blog and opinion article discussions on climate change.</p> <p>Selecting Databases and Sources</p> <p>Major Italian News Websites:</p> <p>La Repubblica - Environment and Climate section. Il Sole 24 Ore - Environment and Climate section. Corriere della Sera - Environment section.</p> <p>Popular Italian Environmental Blogs:</p> <p>GreenMe - Blog focused on sustainability and environment. Eco dalle Città - News and information on the environment and sustainability. Il Blog di Beppe Grillo - Section dedicated to the environment and climate.</p> <p>Academic Platforms and Independent Blogs:</p> <p>The Conversation Italia - Articles written by academics and researchers. Medium Italia - Various influential people and experts write in-depth pieces on climate change.</p> <p>Steps to Collect Data</p> <p>Define Keywords and Hashtags: Keywords: "cambiamento climatico", "riscaldamento globale", "azione climatica", "politica ambientale".</p> <p>Web Scraping Tools:</p> <p>Python Libraries: BeautifulSoup and Scrapy for web scraping. APIs: Check if the chosen platforms have APIs for easier data access.</p>

Table 3

ChatGPT Gives Assistance on Sampling, Ethical Considerations, and Analysis

<p>Prompt: Tell me more about Case 4 on sampling, ethical considerations, and analysis (qualitative and quantitative).</p> <p>Response: Certainly! Let's delve deeper into Case 4, which involves analyzing blog and opinion article discussions on climate change. We will cover sampling strategies, ethical considerations, and both qualitative and quantitative analysis methods.</p> <p>Case 4: Analysis of Blog and Opinion Article Discussions on Climate Change</p> <p>1. Sampling Strategy</p> <p>Goal: Ensure a representative and diverse sample of articles to capture a wide range of themes and opinions.</p> <p>Steps:</p> <p>Define the Population: Identify all potential sources, including major news websites, environmental blogs, and academic platforms.</p> <p>Stratified Sampling: Divide the population into different strata (e.g., mainstream news, independent blogs, academic articles) and ensure proportional representation from each stratum.</p> <p>Time Period: Collect articles over a consistent time frame (e.g., January 1, 2022 to December 31, 2023) to observe trends and changes over time.</p> <p>Keywords and Filters: Use keywords such as "climate change", "global warming", "climate action", and "environmental policy" to filter relevant articles.</p> <p>Random Sampling: Within each stratum, randomly select articles to avoid bias. Aim for a total sample size of 500-1,000 articles.</p>

Example:

From major news websites (e.g., La Repubblica, Il Sole 24 Ore), aim to collect 200 articles.

From environmental blogs (e.g., GreenMe, Eco dalle Città), aim to collect 200 articles.

From academic and independent platforms (e.g., The Conversation Italia, Medium Italia), aim to collect 200 articles.

2. Ethical Considerations**Privacy and Data Protection:**

Anonymity: Ensure that any personal information about authors or commenters is anonymized.

Informed Consent: Although difficult with publicly available articles, clearly state the research purpose when reaching out to any authors for clarification or further information.

Data Storage: Securely store collected data to prevent unauthorized access. Use encryption if necessary.

Respect for Source Content:

Fair Use: Ensure that the use of article excerpts complies with fair use policies. Avoid extensive reproduction of content without permission.

Citation: Properly cite all sources to give credit to original authors and publications.

Minimizing Harm:

Sensitive Topics: Be mindful of the potential impact of discussing climate change, especially when dealing with vulnerable populations or controversial opinions.

Balanced Representation: Aim to present findings in a balanced manner, avoiding sensationalism or bias.

3. Analysis Methods**Qualitative Analysis:****Thematic Analysis:**

Coding: Identify and code key themes and patterns in the articles. Use software like NVivo or ATLAS.ti for efficient coding.

Theme Development: Group similar codes into broader themes (e.g., political responses, public awareness, climate activism).

Interpretation: Analyze the themes to understand underlying meanings and implications.

Content Analysis:

Frequency Analysis: Count the frequency of specific terms or phrases related to climate change.

Contextual Analysis: Examine the context in which these terms are used to understand sentiment and framing.

Quantitative Analysis:**Sentiment Analysis:**

Use natural language processing (NLP) tools to analyze the sentiment of articles (positive, negative, neutral).

Tools like VADER, TextBlob, or commercial APIs (e.g., IBM Watson, Google Cloud Natural Language) can be used.

Statistical Analysis:

Trend Analysis: Examine how themes and sentiments change over time using time-series analysis.

Correlation Analysis: Identify correlations between different variables (e.g., source type, sentiment, publication date).

Visualization:

Graphs and Charts: Use bar charts, line graphs, and pie charts to visualize the distribution of themes and sentiments.

Heat Maps: Create heat maps to show the intensity of discussions on specific topics over time.

Subsequently, I extracted an article from the online news platform of “La Repubblica”, following ChatGPT’s instructions¹. At that point, I asked ChatGPT to perform a qualitative thematic analysis of the article (I copied and pasted the text of the article after the prompt) (Table 4). ChatGPT broke down the content into major themes such as the environmental impact of climate change, human responsibility through daily actions, and the role of technological innovation and corporate responsibility. It highlighted specific sub-themes, including the rapid melting of the Arctic, commitment to sustainability, and collaborative efforts with National

¹ The article can be found at the following web address https://www.repubblica.it/native/ambiente/2021/09/15/news/riscaldamento_globale_evitare_il_punto_di_non_ritorno_dipende_dalle_nostre_scelte-317258056/?ref=drla-2.

Geographic (Table 4). ChatGPT continued with ethical considerations for presenting the data and suggested integrating qualitative thematic analysis with a quantitative analysis of the article. Specifically, ChatGPT suggested developing heat maps to illustrate the intensity of discussions around specific sub-themes over time. When I asked to create the heat maps, it provided step-by-step Python syntax.

Table 4

ChatGPT Performs Qualitative Thematic Analysis

<p>Prompt: Perform a qualitative thematic analysis of this article I found in La Repubblica: ‘Global Warming, Avoiding the Point of No Return Depends on Our Choices’ [Following this, I pasted the full text of the online article].</p> <p>Response: Thematic Qualitative Analysis of the Article Title: “Global warming: avoiding the point of no return depends on our choices” Publication Date: September 15, 2021 Source: La Repubblica</p> <hr/> <p>Main Themes Identified: Climate Change and Environmental Impact: Sub-theme: Melting Arctic and Global Consequences: The article emphasizes the rapid melting of the Arctic, which is occurring at twice the rate of the rest of the planet. This has devastating repercussions, such as the release of methane, a potent greenhouse gas, which accelerates global warming in a destructive feedback loop.</p> <p>Sub-theme: Global Phenomena and Disasters: The article connects climate change to various natural disasters, including wildfires in Siberia, rising sea levels, and extreme weather events like the Texas snowstorms in February 2021. These are presented as direct consequences of climate change, highlighting the urgent need for action.</p> <p>Human Responsibility and Daily Actions: Sub-theme: Importance of Daily Behaviors: The article stresses that individual daily actions are crucial in combating climate change. It implies that small, everyday choices can collectively lead to significant environmental benefits.</p> <p>Technological Innovation and Corporate Responsibility: Sub-theme: (...) Commitment to Sustainability: The article showcases Epson, a Japanese technology company, as a leader in environmental responsibility. It highlights their development of eco-friendly inkjet printers that significantly reduce CO₂ emissions compared to traditional laser printers.</p> <p>Sub-theme: Corporate Milestones and Future Goals: Epson’s milestones, such as achieving zero emissions in recycling by 2004 and the goal of using 100% renewable energy by 2023, are detailed. The company’s future commitments, including substantial investments in decarbonization and becoming carbon negative by 2050, are also discussed.</p> <p>Collaborative Efforts and Partnerships: Sub-theme: Collaboration with National Geographic: The article concludes by mentioning collaboration with National Geographic to raise awareness and reduce environmental impact. This partnership is portrayed as a strategic move to further environmental education and action. (...)</p>

I asked ChatGPT to create an image that could synthesize the themes of the article I selected. ChatGPT’s response was effective in outlining a visual representation of the article’s themes (Figure 1).

Prompt

Can you create an image that summarizes the article?

Response

(...) I will proceed to create this image for you.



Here is the synthesized image that summarizes the key points of the article on climate change from La Repubblica. The image includes:

Title and Date: At the top, highlighting the main focus.

Arctic Melting: On the left, illustrating the melting ice and its global impact.

Natural Disasters: In the center, with icons representing wildfires, floods, and snowstorms.

Daily Eco-friendly Actions: Illustrations of recycling and using energy-efficient products.

Technological Innovations: On the right, showing inkjet printers and their environmental benefits.

Environmental Milestones: A timeline of achievements and future goals.

National Geographic Collaboration: Indicating the partnership for environmental awareness.

This visual representation encapsulates the main themes and messages from the article.

Figure 1. ChatGPT creates an image that synthesizes the themes of the article.

To test its descriptive quantitative analysis capabilities, I asked ChatGPT to count the most significant words recurring in the selected article (excluding conjunctions). ChatGPT provided me with a bar chart focusing on the top 5 significant words in the climate change article (Figure 2).

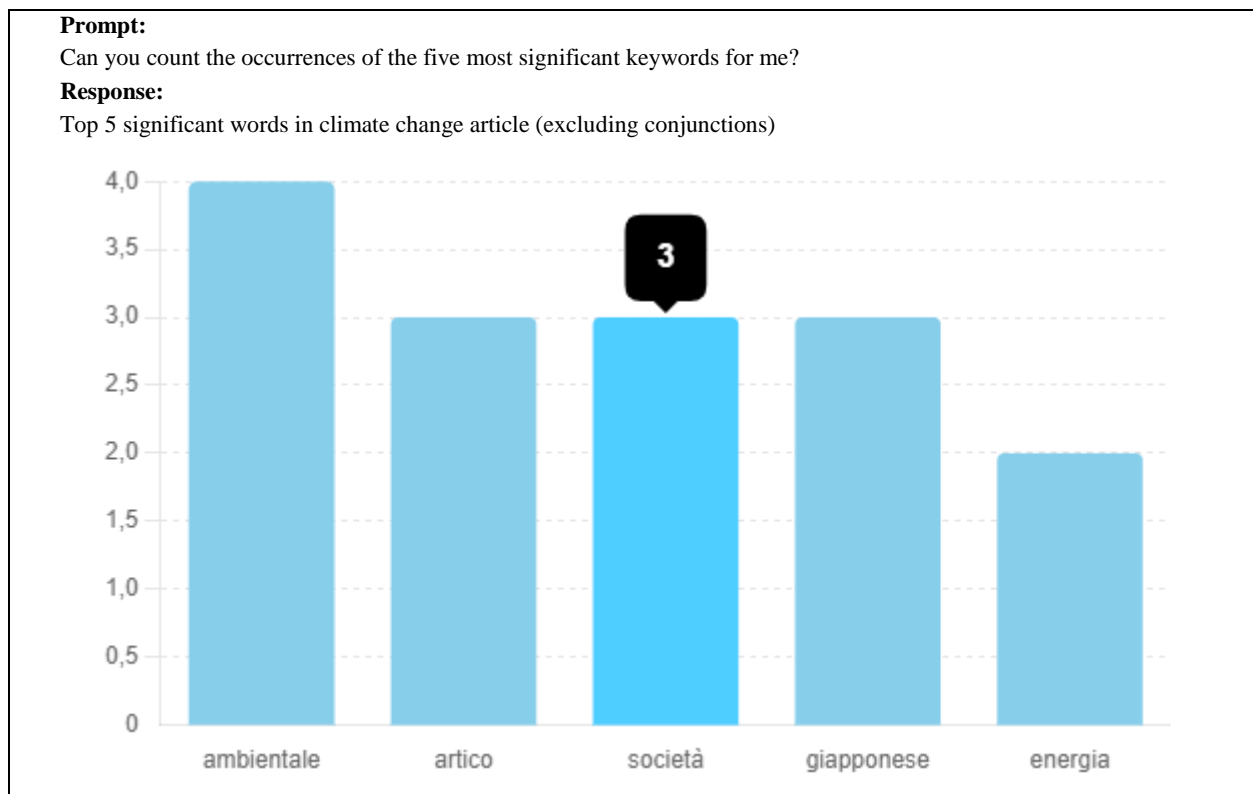


Figure 2. ChatGPT provides a bar chart that focuses on the top 5 significant words in the climate change article.

To ensure a good presentation to the scientific community, I turned to ChatGPT for ideas and strategies. The insights are summarized in Table 5, where I have outlined the most valuable recommendations.

Table 5

ChatGPT Offers Ideas to Ensure a Good Presentation to the Scientific Community

<p>Prompt: How could I present this study to the scientific community?</p> <p>Response: Presenting your study to the scientific community involves several key steps to ensure your findings are communicated effectively and credibly. Here's a comprehensive guide on how to present your study:</p> <p>1. Prepare a Detailed Research Paper</p> <p>Structure:</p> <p>Title: Clear and concise, indicating the main focus of your study.</p> <p>Abstract: A brief summary of the study, including objectives, methodology, key findings, and conclusions.</p> <p>Introduction: Introduce the topic, outline the research question, and state the significance of the study.</p> <p>Literature Review: Review existing research on climate change, particularly focusing on web-mediated document analysis.</p> <p>Methodology: Describe your sampling strategy, data collection methods, ethical considerations, and analysis techniques (both qualitative and quantitative).</p> <p>Results: Present your findings with appropriate charts, graphs, and heat maps.</p> <p>Discussion: Interpret the results, discuss their implications, and compare them with existing literature.</p>

Conclusion: Summarize the key points, state the limitations of the study, and suggest areas for future research.

References: Cite all sources used in your research.

2. Create a Presentation

Slides Structure:

Title Slide: Title of your study, your name, and affiliation.

Introduction Slide: Brief overview of the study.

Literature Review Slide: Key points from the literature review.

Methodology Slide: Overview of your sampling, data collection, and analysis methods.

Results Slides: Visual representation of your findings using graphs, charts, and heat maps.

Discussion Slide: Key insights and their implications.

Conclusion Slide: Summary of findings and future research directions.

Q&A Slide: Invite questions and discussion.

Tips:

Use bullet points for clarity.

Include visuals to make your presentation engaging.

Keep text to a minimum and focus on key points.

3. Submit to Academic Journals

Journal Selection:

Choose journals that focus on environmental science, climate change, or digital media analysis.

Ensure the journal's scope matches your study's focus.

Submission Process:

Follow the journal's submission guidelines carefully.

Prepare a cover letter explaining the significance of your study and why it's suitable for the journal.

4. Present at Conferences

Conference Selection:

Look for conferences on climate change, environmental studies, or digital humanities.

Check the call for papers and submission deadlines.

Preparation:

Prepare an abstract and submit it to the conference.

If accepted, prepare a presentation as outlined above.

Practice your presentation to ensure clarity and timing.

5. Engage with the Academic Community Online

Platforms:

Share your research on academic platforms like ResearchGate or Academia.edu.

Join relevant forums and groups on LinkedIn and other social media platforms to discuss your findings.

Engagement:

Publish blog posts summarizing your research on platforms like Medium or your institution's blog.

Engage with other researchers by commenting on related studies and participating in online discussions.

Example Outline for a Conference Presentation

[Chat presents the structure of a 10-slide presentation, from the title to the closing remarks]

(...)

By following these steps, you can effectively present your study to the scientific community, ensuring your research is well-received and contributes to the broader discourse on climate change and web-mediated document analysis. If you need further assistance or specific examples, feel free to ask!

I could have asked for help with the bibliography, either in terms of citations or at least to improve the formatting standard. However, I decided to end the session. Although it might not be the norm, I took a moment to thank ChatGPT for the work done and say goodbye (Table 6).

Table 6

I Thank ChatGPT for the Work Done and Say Goodbye

<p>Prompt: We have worked a lot and well, thank you and goodbye.</p> <p>Response: You're welcome! It has been a pleasure assisting you. If you have any more questions in the future, don't hesitate to reach out. Goodbye and good luck with your research!</p>
--

ChatGPT, the Others and New Horizons

The experience of integrating ChatGPT-4 into my research workflow over an intense four-hour session has revealed both the vast potential and the nuanced role of AI in the research process.

On initial review, ChatGPT provides coherent responses and generates content that aligns with the relevant methodological literature. However, it does not appear to be highly original in the material it produces, as it tends to propose well-established procedures and methods of analysis. This is certainly a positive aspect, but it also somewhat limits its potential. It would be interesting to see how well-crafted and targeted prompts can encourage more creative responses, pushing ChatGPT to explore unconventional solutions. Further research will be crucial to compare the quality of work produced by human researchers with that generated by artificial intelligence in relation to characteristics such as accuracy, creativity, speed, flexibility, coherence, interdisciplinary integration, and ethics.

Throughout the session, ChatGPT provided a range of ideas and suggestions, sometimes giving me the impression of collaborating with another researcher, and at other times acting as a highly efficient assistant, ready to execute tasks as directed. This dual role highlighted the importance of my own input—my questions, subjective decisions, and strategic choices—which ultimately shaped the research outcomes. AI is not a tool that automatically creates projects or processes and analyzes data. My impression is that the integration of AI—specifically ChatGPT—requires clear instructions and active participation by researchers.

While ChatGPT proved to be a valuable support, particularly in the initial stages of theme generation and development, it was clear that the deeper, more interpretive aspects of research remain the domain of human expertise. The AI was adept at offering operational assistance, but the responsibility for making sense of the data and constructing a meaningful narrative fell to me. This aligns with the broader understanding that while AI can automate and enhance certain research tasks, the creation of meaning, contextual understanding, and ethical considerations are areas where human insight is indispensable.

This study further illustrates the effective collaboration between human researchers and AI, merging the capabilities of tools like ChatGPT with the critical thinking and contextual awareness unique to humans. This hybrid approach not only suggests the feasibility of AI as a research assistant but also opens up new possibilities for its application in academic research. The concept of “hybrid teams”, where humans and AI work together, highlights the potential for more efficient and innovative research outcomes (see the debate in Dwivedi et al., 2023). However, it raises important issues concerning the transparency of AI operations, data privacy, and the appropriate attribution of AI-generated content. A lack of clarity in these areas could undermine trust in the results, compromise privacy, and blur the line between AI-generated and human work.

It is important to recognize that while ChatGPT-4 was used in this study, numerous AI tools are available, each with its own unique strengths. Additionally, AI tools are continuously evolving, expanding their

capabilities. Given this rapid advancement in technology, focusing on the technical details of individual programs is of limited value. Instead, the emphasis should be on the strategic application of AI tools to enhance the overall research process—learning how to design effective prompts, interpret AI outputs, and seamlessly integrate these insights into broader research frameworks.

References

- Arosio, L. (2010). Personal documents on the Internet: What's new and what's old. *Journal of Comparative Research in Anthropology and Sociology*, 1/2, 1-16.
- Arosio, L. (2022). What people leave behind online: Digital traces and web-mediated documents for social research. In F. Comunello, F. Martire, and L. Sabetta (Eds.), *What people leave behind: Marks, traces, footprints and their relevance to knowledge society* (pp. 311-323). Springer Nature.
- Banker, S., Chatterjee, P., Mishra, H., & Mishra, A. (2023). Machine-assisted social psychology hypothesis generation. PsyArXiv.
- Chen, T.-J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. *Journal of the Chinese Medical Association*, 86(4), 351-353.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.
- Grossmann, I., Feinberg, M., Parker, D., Christakis, N., Tetlock, P., & Cunningham, W. (2023). AI and the transformation of social science research. *Science*, 380(6650), 1108-1109.
- McCulloch, G. (2004). *Documentary research: In education, history and the social sciences*. London and New York: Routledge.
- McGee, R. W. (2023). Using ChatGPT to conduct literature searches: A case study. *Journal of Business Ethics*, 95(2), 165-178.
- OpenAI. (2024). *Prompt engineering guidelines*. Retrieved from <https://platform.openai.com/docs/guides/prompt-engineering>
- Park, Y. J., Kaplan, D., Ren, Z., Hsu, C.-W., Li, C., Xu, H., ... Li, J. (2024). Can ChatGPT be used to generate scientific hypotheses? *Journal of Materiomics*, 10(3), 578-584.
- Plummer, K. (1983). *Documents of life*. London: Allen & Unwin.
- Prior, L. (2003). *Using documents in social research*. SAGE Publications.
- Scott, J. (1990). *A matter of record: Documentary sources in social research*. Polity Press.
- Xu, R., Sun, Y., Ren, M., Guo, S., Pan, R., Lin, H., ... Han, X. (2024). AI for social science and social science of AI: A survey. *Information Processing & Management*, 61(3), 103665.
- Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2023). Can large language models transform computational social science? arXiv:2305.03514.