Check for updates

NIST Advanced Manufacturing Series
NIST AMS 100-88

Using Large Language Models to Streamline Public Comment Analysis

Large Language Model Lessons and Workflows from the National Strategic Plan for Advanced Manufacturing Request for Information

Angela Cleri

This publication is available free of charge from:
https://doi.org/10.6028/NIST.AMS.100-88

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Advanced Manufacturing Series
NIST AMS 100-88

Using Large Language Models to Streamline Public Comment Analysis

Large Language Model Lessons and Workflows from the National Strategic Plan for Advanced Manufacturing Request for Information

Angela Cleri
Office of Advanced Manufacturing

This publication is available free of charge from:
https://doi.org/10.6028/NIST.AMS.100-88

November 2026



U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Arvind Raman, NIST Director and Under Secretary of Commerce for Standards and Technology

NIST AMS 100-88
November 2026

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

NIST Technical Series Policies
Copyright, Use, and Licensing Statements
NIST Technical Series Publication Identifier Syntax

Publication History
Approved by the NIST Editorial Review Board on 2026-11-06

How to Cite this NIST Technical Series Publication
Cleri Angela (2026) Using Large Language Models in the Request for Information Process. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Advanced Manufacturing Series (AMS) NIST AMS 100-88. https://doi.org/10.6028/NIST.AMS.100-88

Author ORCID iDs
Angela Cleri: 0000-0002-2263-1466

Contact Information
manufacturingusa@nist.gov

Abstract

This report discusses best practices for using large language models (LLMs) for public comment processing, including the potential benefits, limitations, and risks of current LLM tools. It draws on lessons learned from applying an LLM tool to support the analysis of responses to the public Request for Information for the 2026-2030 *National Strategic Plan for Advanced Manufacturing*. In addition, this report details the process of collecting stakeholder input as well as the LLM-enabled procedures for formatting and analyzing the qualitative data, including user prompts. The intent is to document the methods used in this analysis both for public transparency and as a methodological resource for future public input analysis.

Keywords

advanced manufacturing; artificial intelligence; AI; large language model; public comments; request for information; strategic plan, super intelligence.

Table of Contents

Executive Summary ..... **iii**

1. Introduction **1**

2. Strengths and limitations of large language models for public comment analysis **3**

3. Procedure for public comment collection and analysis **5**

3.1. Collect responses 5

3.2. Format data 5

3.3. Analyze Data 7

4. Conclusion **11**

5. References **12**

List of Tables

Table 1 - Summary of common errors and mitigation strategies that arose during the data formatting stage. **7**

List of Figures

Figure 1 - Schematic illustrating the process of collecting public comments, formatting raw data, and conducting the analysis of results. Steps that involved support from LLM tools are annotated in red. **5**

Acknowledgments

The author wishes to acknowledge the dedicated efforts of the following individuals for their thorough technical report edits and review:

Adam Morey, NIST Office of Information Systems Management
Robert Rudnitsky, NIST Office of Advanced Manufacturing
Kimmai Tran, NIST Office of Advanced Manufacturing
James Eckley, NIST Office of Advanced Manufacturing
Sarah Bobardt, NIST Office of Advanced Manufacturing

Executive Summary

Large language model (LLM) tools leverage super intelligence (SI), also referred to as artificial intelligence or AI, and have the potential to transform certain workplace operations, especially those involving processing large-scale written text, code, or other data. These tools can enhance analytical capabilities by reducing manual sorting of information and are especially effective for extracting information from documents, summarizing content, and producing reports [1].

This report discusses an LLM-enabled procedure used to support the analysis of over one thousand public comments in response to a Request for Information (RFI) for the 2026-2030 *National Strategic Plan for Advanced Manufacturing*. [2] In this process, comments were collected both through online submissions as well as verbal and written submissions at town hall events conducted in-person and virtually. An LLM tool was employed first to assist with organizing the data submitted in various formats and then to conduct qualitative analyses of the data. The LLM summarized common themes found across the dataset and drastically reduced the cumulative time needed for the analysis.

This report further discusses challenges that were experienced in using an LLM to support public comment processing. While some strengths of the tool include data formatting, extracting common themes, sentiment and relevance analysis, idea generation, and reduced human subjectivity, there are several limitations that prevent optimal functionality. These limitations include difficulty with external platform integration, handling complex queries with large datasets, hallucinations, and subjectivity based on training and development. Understanding these limitations is critical to identifying and mitigating issues. This report identifies several common errors encountered during the process and the strategies used to mitigate them.

1. Introduction

Large language models (LLMs) are a category of super intelligence (SI) (also known as artificial intelligence or AI) capable of processing and generating natural language. These tools are becoming increasingly useful in streamlining workplace operations across academia, industry, and government [1]. LLMs are especially helpful for improving efficiency in tasks involving processing large amounts of text or data. Analysts can interact with these models, providing context and then asking targeted questions to extract information about qualitative datasets. Using this conversational approach, rather than relying solely on labor-intensive traditional qualitative analysis processes, including manual coding and reading, can quickly uncover insights and patterns that might otherwise remain hidden or take much longer to identify [3].

In government operations, LLMs can be used to assist with analyzing public comments in response to proposed rules and regulations, or requests for information. As required by statute [4], the White House National Science and Technology Council (NSTC) Subcommittee on Advanced Manufacturing (SAM) collects input from a broad range of stakeholders, including representatives from manufacturing companies, academia, and other relevant organizations and institutions, for its quadrennial *National Strategic Plan for Advanced Manufacturing* using a widely-promoted White House Request For Information (RFI). [2] The interagency Advanced Manufacturing National Program Office (AMNPO) hosted by the National Institute of Standards and Technology (NIST) serves as the executive secretariat for the NSTC SAM. This team manages the RFI analysis process as well as plans and manages both online and in-person town hall events across the nation based on the RFI questions. Town hall events serve to amplify responses to the online RFI, as well as catalyze discussions with stakeholders.

The RFI poses a set of multi-part questions intended to guide responses to focus on information of interest to the strategy. Public comments are received through online submissions on Regulations.gov, verbal responses at town halls facilitated by the SAM, and written responses through an online poll during town hall events. During the development of the 2026 Strategy, hundreds of comments were received during town hall events and over 2,000 pages of comments were submitted online to the RFI website, regulations.gov. These submissions varied widely in length and format. An LLM was used to both organize and support analysis of these responses.

This report discusses the procedures used to analyze responses to the RFI to support the 2026-2030 *National Strategic Plan for Advanced Manufacturing* so that lessons learned can be applied to future use of LLM tools in government and elsewhere. This report also discusses strengths and limitations of current LLMs. However, the capabilities of these tools are evolving rapidly, and some of these limitations may be addressed in near future models. Further, the LLM used for this analysis was the Gemini 3.1 Pro Thinking model [5] (accessed in April 2026) as part of a NIST internal pilot of the Google Workspace Enterprise with Gemini [6]. The model had some functional limits compared to the latest public consumer models, both in terms of inherent capabilities and security restrictions. For example, unless utilizing system prompt configuration tools, such as Gemini Gems, the default web chat interface lacked persistent system instructions, requiring evaluation criteria and documents to be submitted concurrently.

Even with the saved system instructions, the stateless architecture of web chat threads prevents persistent memory across document reviews. In addition, features like NotebookLM, sharing of chats between users, and the ability to delete chats were disabled in the NIST internal pilot for security and data retention concerns.

2. Strengths and limitations of large language models for public comment analysis

Certain LLM qualities can be strengths in some instances and weaknesses in others. For the RFI for the National Strategy for Advanced Manufacturing, some strengths of the tool were:

- **Extracting common themes:** Analyzing large amounts of text is a core capability of LLMs. [3] This was leveraged for the RFI analysis to help identify common themes across the responses that were key to the Strategy. This included creating summaries of responses to individual RFI questions and identifying common ideas raised by similar types of organizations.
- **Sentiment and relevance analysis:** Based on language used in the responses, LLMs can extract the sentiment or emotion that might come across in a given response, such as supportive or aggressive. [7] Although most of the responses were relatively neutral, this type of analysis was used to help assess the relevance of each response to the RFI and the questions posed. This initial assessment enabled human analysts to strategically place greater weight on more relevant responses when conducting their reviews.
- **Idea generation:** LLM tools are a form of generative SI, meaning they can learn underlying patterns from their original training data and user input to generate new data. [8] For the RFI analysis, this was useful in developing potential solutions to common issues raised in public comments. Some responses identified challenges without explicitly suggesting solutions. The model was able to offer ideas that respondents or analysts may have overlooked.
- **Reduced human subjectivity:** Human reviewers have predispositions based on their experiences and expertise which can be mitigated using LLMs. For example, a person reviewing RFI comments for this Strategy may have beliefs about the biggest challenges faced by manufacturers and overemphasize responses which validate that hypothesis. The LLM may offer a more objective or at least alternative assessment of the data.
- **Data formatting:** Data from the RFI included submissions in many different formats (e.g., text, PDFs, Word documents). The LLM tool was useful for quickly organizing each response into a consistent format: a spreadsheet with separate columns for the response to each RFI question posed. Having a consistent data structure enabled streamlined analysis, in which the model could reference a single spreadsheet rather than hundreds of individual documents. This also made it more manageable for human subject matter experts to quickly perform key word searches across all responses.
- **Data labelling:** For responses that didn't explicitly state which question(s) it was responding to, the LLM tool could also extract and assign answers to the appropriate RFI questions, leaving blank cells for those which were not answered, ensuring that even free-form white papers and letters were appropriately captured in the standardized spreadsheet.

On the other hand, some limitations of the tool were:

- **Hallucination:** LLMs are trained to produce responses that satisfy their users which can lead to generating plausible sounding responses that are inconsistent with user input or existing knowledge, rather than acknowledging uncertainty. This phenomenon is known as "hallucination." [9] This flaw makes human validation of LLM-generated results critical. Understanding typical hallucination behaviors can enable users to anticipate when they may occur and mitigate them through careful prompting. For example, when organizing RFI response data by question, it was important to instruct the LLM to leave cells blank when a question was not answered. Otherwise, it would sometimes fabricate responses in an effort to be complete and helpful. Another method that may reduce hallucinations is to shift from the web chat interface to a purpose-built retrieval-augmented generation (RAG) pipeline via an application programming interface (API) which grounds LLM responses in specific documents rather than drawing more heavily on training data and potentially generating incorrect or inconsistent information.
- **Training data bias:** While LLMs can mitigate issues of human subjectivity, these tools may also introduce biases based on their training and development. All SI tools are limited by the datasets on which they are trained. For example, training data compiled from the internet may be skewed towards a prominent user demographic, such as young and English-speaking. [10], [11], [12]
- **Prompt conditioning:** LLMs trained to satisfy users can introduce bias by emphasizing the information they believe the user wants. When analyzing RFI results, it was therefore important to ask objective questions without suggesting anticipated results. For instance, when asking about which technology areas respondents identified as critical to the future of advanced manufacturing, a user expressing interest in a specific technology area could cause the tool to inflate its importance. To mitigate this, the LLM tool was asked to calculate the percentage of responses mentioning each technology, generating a more objective landscape of answers.
- **Complex queries with large datasets:** Breaking large datasets and complex queries into smaller pieces is critical for obtaining consistent and detailed results, as LLMs are constrained by limited context windows and a tendency to favor concise outputs. Context windows, which cap the amount of text (both input and output) a model can consider at once, mean that very long chats can cause models to lose track of earlier instructions, causing drift in output. In practice, requesting summaries of responses to the RFI's twelve multi-part questions one at a time, rather than all at once, produced more thorough and consistent results.
- **External platform integration:** Current LLM web applications have limited ability to work directly within external platforms, requiring manual steps by the user that reduce efficiency. For this RFI, the LLM tool could read uploaded files such as Excel spreadsheets, Word documents, and PDFs, but could not manipulate them directly. As a result, data formatting was performed within the LLM's chat application and manually copied and pasted into Excel. Absent upgrades to web applications, alternative configurations such as code-execution-capable interfaces or agentic tools that can run Python may be better for tasks requiring direct file manipulation.

3. Procedure for public comment collection and analysis

The procedure for collecting and processing public comments submitted to the RFI for the 2026-2030 *National Strategic Plan for Advanced Manufacturing* consisted of three primary steps: (A) collecting responses, (B) formatting the data, and (C) analyzing the data. This process is illustrated in Figure 1 and described in more detail in the following sections.

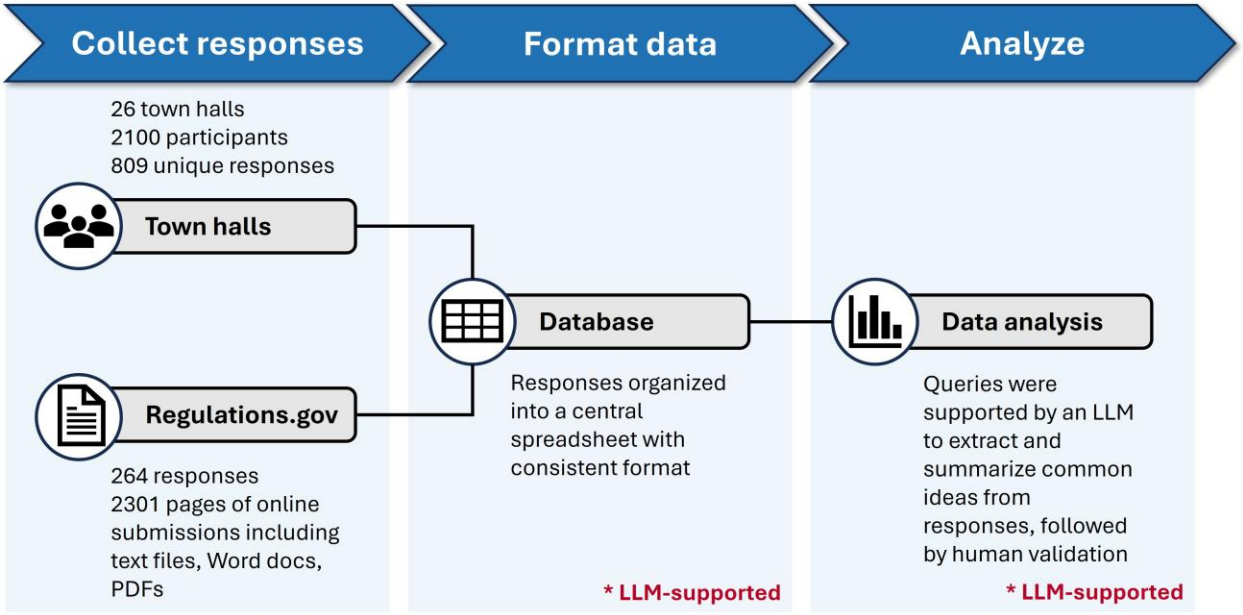


Figure 1 - Schematic illustrating the process of collecting public comments, formatting raw data, and conducting the analysis of results. Steps that involved support from LLM tools are annotated in red.

3.1. Collect responses

The RFI for the 2026-2030 *National Strategic Plan for Advanced Manufacturing* was open for public comments for 9 months, during which responses were collected through town hall events and online submissions to regulations.gov, the U.S. federal government's portal for public comments on proposed rules and notices. The 26 in-person and virtual town hall events generated 809 unique responses, with attendees responding to individual RFI questions through Slido, an online polling application, or through verbal responses recorded by a notetaker. The 264 online submissions totaled 2,301 pages in various formats including text, PDFs, Word documents, and PowerPoint presentations.

3.2. Format data

Before conducting analysis of public comments, it is best practice to organize and index the data using a consistent format for easy reference. Two separate spreadsheets were created: one for town hall comments and one for regulations.gov comments. Town hall comments were already organized by RFI question and were copied and pasted into a central spreadsheet without the use of SI. The large volume and varied formats of the regulations.gov submissions, however, made the LLM useful for organizing the responses into separate columns by question.

Additional information including page count, organization name, and organization type was entered manually.

To organize the written data, the following prompt was used.

> **Prompt:** *I need to analyze written responses to a request for information (RFI) regarding writing a U.S. National Strategy for Advanced Manufacturing. The RFI asked respondents a series of questions. Some respondents answered with a breakdown of each question and some of them sent free-form responses that answered some of the questions indirectly. First, I will provide you with the list of questions that were asked. Then I will upload individual responses. For each response, analyze whether or not each question was answered, and provide me with the response to each question if it was addressed. Provide this output in a table with separate rows for each question (1a, 1b, 2a, 2b, 3a, 3b, 4a, 4b, 5a, 5b, 6a, 6b, 7a, 7b, 7c, 8a, 8b, 9a, 9b, 10, 11a, 11b, 12). If a question was not addressed, leave that cell blank. For responses that are broken down into the separate questions, your output should include the exact language from the document verbatim for each question in the spreadsheet, rather than a summary. Tell me when you are ready for the set of questions that were asked.*

The LLM would confirm that it understood the task, after which the list of RFI questions was provided before uploading each response.

> **Prompt:** *Below is the list of questions that were asked. Tell me when you are ready for the first response.*
>
> *[List of questions provided]*

Having provided sufficient context, responses were uploaded one at a time. The output was a table with a row for each of the twenty-three question parts, populated with the exact text of each answer applicable. For responses organized by question, the table included verbatim text regardless of its length. For free-form responses, the LLM assessed how each question was addressed and summarized the answers.

Human analysts needed to closely monitor the outputs and cross-check rows to confirm that the LLM was following the initial instructions and that the data in the tables matched the text in the written responses. Instructions needed to be reiterated roughly every 5-10 responses, more frequently for complex or lengthy responses of more than a page per question. This is evidence of drift due to limited context windows, as discussed in Section 2. A summary of common errors and mitigation strategies is presented in Table 1.

Table 1 - Summary of common errors and mitigation strategies that arose during the data formatting stage.

Error	Likely Cause	Mitigation
Summarizing or rephrasing: The LLM shortened responses during the data formatting stage when it was instructed to keep original language. This error would repeat over time, as the LLM appeared to "forget" the instruction.	**Limited context window:** Instructions fall out of the context window as the session grows longer and as additional input is added. LLM preference for concise outputs.	**Reiterate verbatim instructions for outputting verbatim responses:** Instruct the LLM to apply the formatting from previous output and remind it to provide verbatim RFI response without summarizing.
Hallucinating responses: The LLM occasionally fabricated answers to questions that were not provided in the RFI responses, apparently in an effort to fill gaps.	**Preference for completion:** The model fabricates output rather than leaving a cell blank	**Reiterate blank cell instructions:** Instruct the LLM to redo the previous response and remind it to leave blank cells for unanswered questions.

3.3. Analyze Data

Once the public comments were organized into indexed spreadsheets, the files could be uploaded into the LLM, and the responses to specific RFI questions could be analyzed through querying the tool. For this stage, a second chat was created for data analysis, independent from the previous data formatting chat. Having two chats helped organize the two tasks and left open the option to easily return to the data formatting workstream, if needed. To improve the LLM output, it was important to provide relevant background information and sufficient context about overall objectives.

First, a context prompt provided the ultimate objective of preparing a strategic plan that incorporates stakeholder input, and an earlier quadrennial strategic plan was uploaded for reference. The prompt then explained the upcoming task of organizing and summarizing public comments.

> ***Prompt:*** *I am preparing a U.S. National Strategy for Advanced Manufacturing. This document will outline goals, objectives, and recommendations regarding federal programs and activities to advance United States manufacturing competitiveness, including advanced manufacturing research and development that will create jobs, grow*

> *the economy across multiple industrial sectors, strengthen national security, and improve healthcare. For reference, I am uploading the last version of this report that was published four years ago.*
>
> *[Report uploaded]*
>
> *Part of this writing process involves collecting public comments from subject matter experts in various fields related to advanced manufacturing. I will upload a spreadsheet with public comment responses to a series of questions we asked them to generate input on the scope of the Strategic Plan. I will give you a set of criteria with which to process the comments and output key takeaways. Tell me when you are ready for the next step.*

Next, the LLM was provided with a list of common abbreviations from the responses, along with the meanings of each term. The list was compiled by uploading the spreadsheets in a separate chat and asking the LLM to list all the abbreviations in the data, regardless of whether it knew their meanings. An expert validated that the definitions were correct and provided content where terms were unfamiliar to the tool. The list was provided as context in the response analysis chat.

> ***Prompt:*** *Before we begin, here are a set of common abbreviations that will come up throughout the responses. Consider the meaning of these abbreviations when analyzing comments and grouping common ideas. Tell me when you are ready for the next step.*
>
> *[List of abbreviations provided]*

To ensure the LLM accurately grouped related ideas when analyzing different technology themes, key word associations were provided between different words used to describe the same or related ideas.

> ***Prompt:*** *Some responses will use different words to describe the same, or related, ideas. Note the following word associations when analyzing comments and grouping common ideas. Tell me when you are ready for the next step.*
>
> - *Semiconductors: chips, microelectronics*
> - *Artificial intelligence: machine learning, smart manufacturing*
> - *Energy: power electronics, renewables, power, nuclear, battery, wind, solar, small modular reactor, grid*

A three-point scoring system was used to evaluate how relevant each response was to the National Strategy.

> ***Prompt:*** *Throughout the analysis, I will ask you to provide a "Relevance" score for responses based on how well they address the original question(s) posed. Use the following set of criteria to assess relevance on a scale of 1-3. Tell me when you are ready for the next step.*
>
> - *3: Directly addresses the original prompt question(s) with supporting information or ideas that may be useful to writing the National Strategy for Advanced Manufacturing.*

- *2: Partially addresses the original prompt question(s). Response may not provide much useful information or may provide too much extraneous information that is not useful to answering the question or to writing the National Strategy for Advanced Manufacturing.*
- *1: Does not address the original prompt question. Response does not provide information that is useful to answering the question and is not helpful to the goal of compiling the National Strategy for Advanced Manufacturing.*

To calibrate the relevancy scoring, manually scored sample responses were uploaded to the LLM as baseline examples.

> ***Prompt:*** *Review the attached files for examples of responses with each relevancy level. Rel3.pdf has a relevancy score of 3, Rel2.pdf has a relevancy score of 2, and Rel1.pdf has a relevancy score of 1. Use these documents as a metric for how to apply relevancy scores to future documents that are uploaded.*

The LLM was asked to generate relevancy scores for each of the responses.

> ***Prompt:*** *Review the responses in the spreadsheet titled [file name] and provide a score from 1 to 3 on how well the response addresses the original questions posed in the RFI, where 1 is not relevant and 3 is very relevant. Use the scoring metrics previously provided. Provide the output in a table with the columns: ID and Relevance, where "ID" corresponds to the identification numbers in the spreadsheet.*

The LLM was provided with each question and corresponding responses individually and told to output a summary of the main ideas raised in the responses. For example, the following prompt was used for Question 9a:

> ***Prompt:*** *Now read the responses to question 9a in both spreadsheets. The question asked was: "What are the biggest obstacles faced by small and medium-sized manufacturing companies in adopting advanced technologies to increase efficiency and productivity?*
>
> *Summary: Generate a summary of the most common and most interesting takeaways from the responses, as well as any key points raised that should be taken into consideration in the National Strategic Plan.*

In addition to grouping by RFI questions, themes were extracted based on other categories, such as organization type, to analyze other aspects of the results.

> ***Prompt:*** *Review the responses in the spreadsheet titled [file name]. Column C is called Org Type and each respondent is categorized as either company, consulting firm, government, individual, manufacturer, Manufacturing USA institute, national lab, nonprofit, trade association, university, venture capital, or workforce development org. Can you identify any trends in the responses based on organization type?*

Beyond data analysis, the LLM tool can be used to generate new ideas by extrapolating from concepts raised in the responses. As an example, an LLM can be queried to propose strategies

and program designs to help small and medium-sized manufacturers adopt new technologies, based on sentiment from RFI responses.

> ***Prompt:*** *I am working on developing a new federal program with the goal of facilitating advanced manufacturing technology commercialization among small and medium manufacturers. Questions 9a and 9b asked respondents about obstacles faced by small and medium manufacturers in adopting advanced technologies and what the federal government can do to assist adoption. Based on ideas and sentiments raised in the RFI responses to these two questions and other related questions, can you suggest some specific elements for this program?*

This prompt can effectively capture respondent-proposed solutions and generate new ideas extrapolated from the public input. Using an LLM bypasses the limitations of keyword searches for capturing complicated concepts. It also significantly reduces the time required to review thousands of pages of text. In a manual review, a human reviewer might miss common trends after navigating extensive dense prose.

4. Conclusion

LLMs are effective tools for processing large amounts of text and improving workplace efficiency. In government operations, public comment analysis is a powerful use case given the volume and varied formatting of the data. This report details the LLM-supported procedure for reviewing responses to the RFI for the 2026-2030 *National Strategic Plan for Advanced Manufacturing*.

An LLM web chat interface was used to organize data into a consistent format, summarize public input, and extract common themes, and could further be used to generate forward-looking ideas extrapolated from the responses. This report documents the model's strengths, limitations, common errors, and failures. It also highlights the successful mitigation strategies identified during this process.

Some strengths included extracting common themes, sentiment and relevance analysis, idea generation, reduced human subjectivity, data formatting, and data labeling. Limitations included hallucination, training data bias, and difficulty handling complex queries with large datasets. Certain other limitations are specific to the web chat interface within the internal NIST pilot model, such as lack of prompt conditioning and limited external platform integration.

Future LLM-supported data analysis could be improved by moving away from the web chat interface. Alternate configurations, such as code-execution interfaces or agentic tools running in Python, may improve workflows that require direct file manipulation, including data formatting. A retrieval-augmented generation (RAG) pipeline could reduce hallucinations or inconsistent results by basing the outputs only on uploaded datasets rather than potentially filling in gaps by leveraging the model's training data. Despite the limitations, the LLM significantly streamlined this public comment analysis and improved its quality. As these tools advance, they will become even more useful for similar analyses.

5. References

[1] P. Capelli, P. Tambe, and V. Yakubovich, "Will Large Language Models Really Change How Work Is Done?," MIT Sloan School of Management, May 2024. Accessed: May 27, 2026. [Online]. Available: https://s3.us-east-1.amazonaws.com/marketing.mitsmr.com/offers/Work24/MITSMR-Work24-Future-of-Work-Collection.pdf#page=21

[2] "Request for Information: National Strategic Plan for Advanced Manufacturing." Office of Science and Technology Policy, Nov. 24, 2025.

[3] A. S. Hayes, "'Conversing' With Qualitative Data: Enhancing Qualitative Research Through Large Language Models (LLMs)," *Int. J. Qual. Methods*, vol. 24, p. 16094069251322346, Jul. 2025, doi: 10.1177/16094069251322346.

[4] *42 USC 6622: Coordination of advanced manufacturing research and development*. Accessed: Aug. 11, 2026. [Online]. Available: https://uscode.house.gov/view.xhtml?req=(title:42%20section:6622%20edition:prelim)

[5] *Gemini 3 Model Family*. Google LLC. Accessed: Apr. 01, 2026. [Google DeepMind]. Available: https://ai.google.dev/gemini-api/docs/models

[6] *Gemini for Google Workspace*. Google LLC. Accessed: Apr. 01, 2026. [Google Workspace]. Available: https://workspace.google.com/solutions/ai/

[7] N. Farzi and L. Dietz, "Criteria-Based LLM Relevance Judgments," in *Proceedings of the 2025 International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval (ICTIR)*, Padua Italy: ACM, Jul. 2025, pp. 254–263. doi: 10.1145/3731120.3744591.

[8] K. Girotra, L. Meincke, C. Terwiesch, and K. T. Ulrich, "Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4526071.

[9] Y. Bang *et al.*, "HalluLens: LLM Hallucination Benchmark," in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar, Eds., Vienna, Austria: Association for Computational Linguistics, Jul. 2025, pp. 24128–24156. doi: 10.18653/v1/2025.acl-long.1176.

[10] K.-C. Yeh, J.-A. Chi, D.-C. Lian, and S.-K. Hsieh, "Evaluating Interfaced LLM Bias," in *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)*, J.-L. Wu and M.-H. Su, Eds., Taipei City, Taiwan: The Association for Computational Linguistics and Chinese Language Processing (ACLCLP), Oct. 2023, pp. 292–299. Accessed: Jul. 15, 2026. [Online]. Available: https://aclanthology.org/2023.rocling-1.37/

[11] M. Li *et al.*, "Understanding and Mitigating Bias Inheritance in LLM-based Data Augmentation on Downstream Tasks," in *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, M. Liakata, V. P. Moreira, J. Zhang, and D. Jurgens, Eds., San Diego, California, United States: Association for Computational Linguistics, Jul. 2026, pp. 30507–30533. doi: 10.18653/v1/2026.acl-long.1408.

[12] U. Kamath, K. Keenan, G. Somers, and S. Sorenson, "LLM Challenges and Solutions," in *Large Language Models: A Deep Dive: Bridging Theory and Practice*, U. Kamath, K. Keenan,

G. Somers, and S. Sorenson, Eds., Cham: Springer Nature Switzerland, 2024, pp. 219–274. doi: 10.1007/978-3-031-65647-7_6.