UNC SCHOOL OF GOVERNMENT

PUBLIC MANAGEMENT BULLETIN NO. 30 | AUGUST 2025

Using AI to Strengthen the North Carolina Benchmarking Project

Keegan Huynh, Christopher L. Cole II, and Obed Pasha

The North Carolina Benchmarking Project was established in 1995 to help municipalities compare their service and performance trends with other participating units. Each year, throughout August and September, partner municipalities submit performance and management data across eleven services to the University of North Carolina's School of Government. The Benchmarking Project audits and uploads that data on a dashboard and invites participants to eleven performance strategy sessions in November to analyze data and discuss shared challenges and practices likely to resolve those challenges. The information from the strategy sessions and the dashboard is synthesized into annual reports.

Last year, the Benchmarking Project leveraged artifical intelligence (AI) to manage report-development and data-auditing processes. The use of AI in these two tasks significantly reduced the hours spent on them, decreased the error rate, and strengthened outcomes. In this bulletin we lay out how we adopted AI for data auditing and report development. We hope that our experience will serve as an example for others thinking about leveraging AI to improve their processes.

Data Auditing

Historically, the auditing process has been conducted manually by comparing current submissions against prior data to identify anomalies. For instance, a data inputter might mistakenly add an extra zero to a data point or put a decimal in the wrong place, typical errors our team has learned to flag during manual review. While effective, this process is both time-consuming and vulnerable to human error, often requiring multiple rounds of validation.

Keegan Huynh is a special projects coordinator for the North Carolina Association of County Commissioners.

Christopher L. Cole II is a project coordinator for the North Carolina Benchmarking Project.

Obed Pasha is an associate professor of public management at the School of Government. He specializes in performance management and strategic planning.

To streamline this process and improve data integrity, the Benchmarking Project collaborated with the UNC AI Club to automate the identification of outliers. The primary objective of this collaboration was to develop an AI-supported system that automatically flags data points that deviate significantly from historical trends or peer benchmarks. The automation effort proved to be more challenging than anticipated. The dataset spans eleven departments in sixteen municipalities and includes 873 performance metrics collected over the past four years, amounting to 614,592 unique data points. One of the most critical challenges was calibrating the flagging system to strike the right balance between being too restrictive with error parameters and identifying excessive false positives. The system had to be carefully tuned to distinguish between legitimate anomalies and acceptable variance in service performance. Additionally, we wanted to avoid the "black box" problem that often comes with using AI to review datasets, when the users do not fully understand why an error has been flagged.

The UNC AI Club's executable file automatically audits data and explains each flagged data point using a set of categories: z-score, historical-data comparison, peer-municipality comparison, formatting, and missing categories. The automated flagging and accompanying explanations allow the Benchmarking Project team to review flagged data more efficiently. This improvement also enables partner municipalities to understand the reasons behind the flagged data and why it has been sent for review, offering a level of detail that was previously unachievable due to time and capacity constraints.

Audit-process automation enables our team to redirect time and resources toward higher-impact efforts, such as supporting municipalities in interpreting their performance data and translating those insights into actionable improvements.

Looking ahead, the Benchmarking Project plans to continue refining this AI-supported system in partnership with the UNC AI Club. Future iterations may incorporate predictive analytics, allowing municipalities to validate data, anticipate emerging trends, and proactively address service-delivery challenges.

The 2025 Benchmarking Report

The 2025 North Carolina Benchmarking Report marked a significant evolution in how our team reports public-sector performance. This section outlines how we used AI to assist with transcribing, drafting, synthesizing, and validating tasks related to data resulting from the performance strategy sessions.

Two core goals guided the integration of AI into our writing workflow. First, we sought to enhance efficiency and consistency in how we analyzed and wrote about complex qualitative data. Second, we wanted to ensure that the final report retained the same analytical depth, nuance, and accuracy that partners have come to expect.

Before AI became an integral part of our workflow, the N.C. Benchmarking Report was developed through a structured, labor-intensive process that relied heavily on manual synthesis, writing, and review. The writing process began with a detailed synthesis of qualitative data from the performance strategy sessions, which included flip-chart notes, sticky notes, meeting transcripts, and feedback forms. This stage required researchers to read through and manually interpret large volumes of unstructured information to identify themes and performance outcomes relevant to each service area.

Once the themes were identified, research assistants developed outlines for each participating department's section of the report. These outlines served as internal blueprints and required review and approval before the creation of full drafts could begin. From there, researchers wrote and refined multiple drafts, integrating illustrative examples, partner quotes, and cross-referenced strategies. This was a highly iterative process, with multiple rounds of revisions to ensure clarity, accuracy, and cohesion across sections.

While AI tools have been used sparingly in previous years (for simple transcription or summarization tasks), this year represented the first time we embedded AI into nearly every step of the writing process. The integration was designed to mirror our traditional workflow while enhancing it with tools that could process large volumes of qualitative data, generate initial drafts, and maintain a consistent structure and tone across report sections. Below is a step-by-step breakdown of how AI was embedded into the writing process.

Training the AI

The first step was to equip the AI with the background and expectations needed to act as an effective collaborator. This included uploading written instructions and personal notes outlining the goals and structure of the 2025 Benchmarking Report. The notes provided information about the project's purpose, how data for the report was collected, and how the report should be organized. The instructional materials also included previous reports from 2023 and 2024 to serve as style and formatting references. We provided context for those reports with accompanying prompts to ensure that only the form of the previous reports, and not the content, influenced the AI model's training (e.g., "I will now upload some example reports from the past two years. Use these reports as context for guiding the way we write the new report. Please analyze the uploaded example reports to identify key sections, structure, and formatting conventions, level of analysis, detail, tone, and formatting for drafting the 2025 benchmarking report."). These documents helped train the AI to understand the report's tone, analytical depth, formatting conventions, and organizational logic, essentially teaching the model how we write before asking it to generate any content.

Developing the Outlines

Once the AI was trained, we developed outlines for each department-specific chapter. This involved uploading raw materials from the strategy sessions: flip charts, sticky notes, transcriptions, and meeting notes. Then we provided further input with prompts like this one:

> I am working on drafting a chapter for the Emergency Communications department in the Benchmarking 2.0 Report. I need it to follow the same level of detail, structure, and tone as the Police Draft 2 (attached). The first part of this process is to read, analyze, and understand all of the documents attached from the Emergency Communications, PSS 2024 Flip Chart doc, and Transcription. Afterward we will write the chapter step by step. Please note that the strategies and outcomes for Emergency Comms will be different from Police and will use the Flip Chart as an outline. Just read, analyze, and understand the documents. No need to write anything yet, just be prepared and retain the knowledge.

We organized the data by identifying key outcomes and strategies. We grouped similar ideas under thematic umbrellas to create a structured outline and provided headings, subheadings, and descriptions of what information should go in each section.

Each outline followed a standardized format modeled after previously approved examples (like the Police Services section of the Benchmarking Report) to ensure consistency across departments. These outlines served as the backbone of the chapters and were reviewed before drafting began.

Writing the Chapters

With vetted outlines and a trained AI model, we drafted the full chapters of the report. Topic-by-topic outlines were paired with their supporting documents and fed into the AI. We pasted each outlined outcome section into the AI model's chat interface, reuploading the contextual documents (i.e., the meeting transcript and notes) and feeding the AI model an example. We used the Building a Resilient and Well-Staffed Workforce subsection from Police Services section and said,

> I am writing a new section for the report and want it to follow the structure, depth of analysis, and formatting of a prior section that serves as a strong example. To ensure consistency:
> - Read, analyze, and understand all attached documents, including my outline for the new section and the example section.
> - Follow the example section as a guide for structuring, formatting, and analyzing the content in a similar manner.
> - Maintain coherence and alignment with the existing report's tone, analytical approach, and level of detail.
> - Incorporate evidence-based insights from the provided materials, ensuring factual accuracy and consistency with the report's objectives.
> - Now, let's begin drafting the first section based on my outline for the topic of "Quality" below.

Below this we inserted the chapter's topic outline and an example section.

The AI model generated each report section with attention to structure, clarity, and tone. We now had a solid first draft. To refine the Strategies section to include examples from the strategy sessions, we reattached the transcription and session notes and entered the following prompt:

> Great. Now please rewrite the feasible strategies sections such that it is structured as follows. Strategy—Definition. Impact. Example. Follow this example: Ability to Overhire—Hiring additional personnel beyond current vacancies allows agencies to account for anticipated retirements and attrition, ensuring adequate staffing levels and minimizing operational disruptions. This strategy provides greater scheduling flexibility and reduces overtime burdens on existing staff. One department successfully obtained approval to overhire, which enabled it to maintain sufficient staffing levels despite ongoing retirements and resignations. Make sure your stretch goals are formatted properly. See Police Services as an example, but it should be: Topic—Definition. Impact. Challenges Faced.

AI-generated drafts provided clear definitions, analysis of impact, and real-world examples in a format similar to prior reports.

Fact-Checking and Validation

To prevent misinformation or "hallucinated" content from making it into the final report, we developed a separate validation process. Transcripts were reuploaded into a new chat instance, and AI outputs were cross-checked line by line. We pasted our stretch goals into the chat and told the AI model to fact-check each section by comparing it with the transcription. The following prompts were used:

> I recently wrote a report. However, I am unsure if the sections below contain true examples from my recording/transcript of the session.
>
> - Please read, analyze, and understand the transcript. Then compare my examples below and tell me if they are true and whether they actually happened in the transcript (examples were uploaded).
> - Please identify supporting examples in the transcript for each strategy. These examples should be verifiable and true. Do not hallucinate or invent examples or facts that do not exist in the transcript.
> - Now please rewrite the feasible strategies sections such that it is structured as follows. Strategy—Definition. Impact. Example. Follow this example for the way to structure it: Ability to Overhire—Hiring additional personnel beyond current vacancies allows agencies to account for anticipated retirements and attrition, ensuring adequate staffing levels and minimizing operational disruptions. This strategy provides greater scheduling flexibility and reduces overtime burdens on existing staff. One department successfully obtained approval to overhire, which enabled it to maintain sufficient staffing levels despite ongoing retirements and resignations. Great. Please rewrite the strategies sections in paragraph form. Each section should contain three sentences: one sentence on definition, one on impact/significance, and one setting out an an example.

We used search functions to spot-check the fact-checker's accuracy in the transcript. Any unsupported examples were removed or replaced with verified quotes or stories. Each strategy subsection in the report was rewritten using a three-sentence format: (1) a description and definition of the strategy, (2) a description of the strategy's impact, and (3) an example of the strategy's use in practice. This validation process significantly enhanced the report's clarity, accountability, and alignment with real partner experiences.

Refinement and Editing

The final phase of the AI workflow mirrored our traditional editing process, with a few improvements:

- Introductions were revisited and rewritten once all sections were completed to better reflect the content.
- Conclusions were drafted last, ensuring cohesion and flow.
- Section titles were reformatted to be clear, action-oriented, and consistent.

While the AI played a central role in the drafting and synthesizing phases, the human editing process remained essential to refine language, confirm context, and ensure that the report met our quality standards.

Lessons Learned

Integrating AI into the 2025 Benchmarking Report writing process offered several valuable lessons about the strengths and limitations of these tools in a qualitative, narrative-driven context. While the technology enhanced several aspects of the workflow, it also underscored the continued importance of human oversight, critical thinking, and contextual knowledge.

One of the most evident benefits of using AI was the significant improvement in the speed of synthesizing and drafting content. Tasks that previously required hours, such as identifying themes from transcripts or transforming notes into coherent prose, can now be accomplished in a fraction of the time. This allowed researchers to focus more energy on analysis and refinement, rather than on rote transcription or formatting.

AI also played a key role in maintaining structural and tonal consistency across participant department chapters. By referencing prior "gold standard" sections and applying uniform formatting rules, the model helped ensure that each report section followed a predictable, professional format. This contributed to a more cohesive overall product, which was essential given that multiple writers worked on different chapters.

Despite its advantages, the integration of AI was not without challenges. The model occasionally generated plausible-sounding examples or quotes that did not exist in the transcripts. These hallucinations required careful cross-checking to avoid inaccuracies. While AI could surface patterns from data, it sometimes struggled to interpret ambiguous sticky notes, acronyms, short-hand text, or poorly phrased flip-chart content. Human interpretation remained necessary to fill in contextual gaps and ensure that the synthesized themes accurately reflected partner input.

Through this year's process, we identified several best practices for using AI effectively in qualitative public-sector reporting. First, always review the output. AI can generate a strong starting point, but human judgment is essential to validate, edit, and contextualize the text it generates. Second, train the AI with high-quality context. Uploading detailed notes, past examples, and clear writing instructions significantly improves the relevance and accuracy of AI responses. Third, create a dedicated AI chat for verifying examples. Using a separate chat instance for fact-checking helps prevent contamination from prior prompts and ensures a more objective review of transcripts.

Conclusion

Integrating AI into the 2025 N.C. Benchmarking Project transformed our team's approach to qualitative analysis. AI became a partner in the writing process, helping to structure, synthesize, and draft insights drawn from complex municipal discussions. Careful oversight of the AI model's contributions ensured the creation of an accurate and actionable report that remained grounded in the verifiable experiences of its participants.

For public-sector professionals considering the use of AI in their own writing and research, our experience offers a clear takeaway: AI can be a powerful collaborator when used with diligence. With the proper guardrails, particularly contextual training and rigorous validation, AI can enhance the quality and efficiency of public reporting while preserving the depth and nuance that the work demands.