

# *AI as a Budgeting Tool: Panacea or Pandora's Box?*


Michael E. M. Lee<sup>1, c</sup>, Daniel Hayes<sup>2</sup>, Craig S. Maher<sup>3</sup>


Local government officials are consistently tasked with doing more with less. The combination of fiscal institutions, devolution of fiscal federalism, voter frustration, and economic shocks have been the leading causes of these more recent challenges. However, the reality is that smaller local governments, by their very nature, struggle with service provision. This paper explores the next evolution of change to local governments – artificial intelligence (AI), specifically ChatGPT – and its potential for local governments. We discuss the use of AI in budgeting, with a focus on financial data management and statistical analysis, including forecasting and policy recommendations for a small Nebraska municipality. Our experience with ChatGPT highlights its powerful capabilities for data processing. Transitioning to ChatGPT-4 from ChatGPT-3.5 (with or without Python) improved data processing efficiency but introduced financial costs, and we observed that the use of detailed and precise prompts enhances output quality across all versions. Achieving a balance between time, cost, data preparation, and prompt precision is crucial for optimizing the potential of ChatGPT in financial data analysis.


Keywords: Artificial Intelligence, Local Governments, Public Budgeting and Financial Management

The growth of artificial intelligence (AI), particularly ChatGPT, has recently generated a great deal of excitement (and angst). AI offers both opportunities and challenges for policymakers and analysts in public budgeting and finance. For example, a recent study by Valle-Cruz, Gil-Garcia, and Fernandex-Cortez (2020) explored the utilization of AI in public budgeting. However, the methodologies employed in the study are complex, catering mainly to practitioners with advanced training in statistics, coding, and AI applications.

The timing of AI's growth may prove fortuitous for local government professionals who are saddled with "doing more with less" amid labor shortages. This paper adopts an applied

<sup>1</sup>School of Public Administration, University of Nebraska Omaha.  <https://orcid.org/0009-0000-9608-7590>.

<sup>2</sup>School of Public Administration, University of Nebraska Omaha.  <https://orcid.org/0009-0001-3452-2953>.

<sup>3</sup>School of Public Administration, University of Nebraska Omaha.  <https://orcid.org/0000-0001-5562-3473>.

<sup>c</sup>Corresponding Author: [mlee50@unomaha.edu](mailto:mlee50@unomaha.edu).

approach by examining AI applications for local government practitioners. We focus on a small municipality in Nebraska and utilize easily accessible data to explore the data analysis capabilities of three different AI approaches: the free version of ChatGPT, ChatGPT with coding processes from free programming applications like R and Python; and ChatGPT-4, which offers increased data analysis capacity at a monthly subscription fee.

The analyses focused on the same outputs: data description, financial condition indicators, forecasting, and policy recommendations. The goal was to illustrate the strengths and weaknesses of each AI tool and discuss their relevance for policymakers and analysts. Our findings suggest that the free version of ChatGPT has limited potential for data analysis but can serve as a resource for policy guidance. Using ChatGPT as a tutorial for learning Python was also highly effective but proved to be very time-consuming for novices, raising questions about its utility for local government practitioners. Conversely, ChatGPT version 4.0, with a monthly subscription, emerges as a powerful tool for both data analysis and policy guidance.

## Literature Review

From March 2020 to 2022, local government employment reportedly decreased by about 300,000 workers (National League of Cities, 2023). Despite the improvement in hiring in 2023, responses to this labor shortage have been challenging, particularly for employee recruitment and retention (Mission Square, 2023). This challenge is particularly acute in public budgeting and finance. According to Walsh and Nason (2022), “[t]he supply-demand gap for state and local public finance workers is widening faster than the same gap for the finance sector overall, the public sector overall, and the economy overall” (p. 5).

In addition to reexamining recruitment strategies, salaries, and benefits, the use of technology is also a means of helping address some of these personnel challenges. According to Hinkley (2023), “Advanced technologies—algorithms, artificial intelligence, robotic process automation—have begun to change some public jobs significantly, either augmenting or replacing some human decision-making” (p. 6). She notes that the key drivers of technology adoption are efficiency, performance (including reliability and accuracy), transparency, and crisis management (Hinkley, 2023). When we consider some of the analytic functions required of financial personnel - forecasting, comparative analysis, budget analysis, etc. – technology has played a key role. Nevertheless, the adoption of such analytical tools in small local governments is hindered by limited resources, including trained personnel and necessary hardware/software.

Kim and Bretschneider (2004) assert that “[w]hile new IT tools, especially web-based applications, are becoming more and more prevalent in local government, applying new IT continues to depend on the general ability of local governments to obtain, manage and effectively use IT...” (p. 1). What makes this era different from the past 40-50 years of government capacity and fiscal policy discussions is that, for the first time, resources – hardware, software, and personnel – are less demanding for the end user with the advent of AI, and ChatGPT, in particular. What remains to be seen is its value to local policymakers.

A growing body of work explores AI applications in public budgeting, accounting, and financial management. For instance, Anders (2023) highlights ChatGPT’s value for CPAs, identifying resources such as accounting “influencers,” websites, and bloggers. The paper is a useful resource for public finance professionals interested in learning about the array of resources readily available. One of the noted resources is “How to use ChatGPT for Financial

Analysis” (Pop! Automation, 2023), which offers insights into horizontal analysis (recent historical analysis), vertical analysis (e.g., the accounting treatment for gains and losses arising from a cash flow hedge), and financial ratio analysis. Similarly, Larson and Overton’s (2024) new study examines the effectiveness of machine learning for revenue forecasting.

Exploring the technological evolution in public financial management is valuable, given the significant transformations that have made tasks more accessible and efficient. Srivastava and Jain (2023) delineate distinct phases in the Management of Technology (MoT) from 1972 to 1995, emphasizing Information Technology implementation and strategic alignment. Developments from 1996-2010 focused on strategic flexibility and technology learning, while the maturation of MoT from 2011 to 2022 integrated technology with broader strategic objectives. Similarly, Egodawale, Sedera, and Bui (2022) highlight a surge in digital transformation research from 2015, indicating a shift towards sophisticated technological capabilities in public financial management, proposing the RC3O nomological net as a tracing framework. Additionally, Ciancarini, Giancarlo, and Grimaudo (2023) analyzed the digital transformation in public administrations, emphasizing increased demand for digital services, the central role of data, and technology integration in public service delivery, illustrating a shift from a narrow technological approach to a comprehensive and integrated strategy encompassing various aspects of public service and policymaking.

The development of governmental accounting and reporting standards, as advocated by the Governmental Accounting Standards Board (GASB), marks a significant shift in local governments’ fiscal management and transparency. The implementation of GASB Statements like No. 68 and 45 has compelled local governments to disclose pension and post-employment benefit obligations, thereby influencing their fiscal decisions and resource allocations (Coe & Rivenbark, 2010; Dambra, Even-Tov, & Naughton, 2023). Furthermore, GASB’s recent standards, including Statements No. 67 and 68, not only standardize pension reporting, increase transparency, and potentially impact financial statement configurations (Lowe, 2014), but they also address collective action issues, thereby enhancing comparability and assessment in financial conditions across governments (Baber, Beck, and Koester, 2024; Rivenbark & Roenigk, 2011). Moreover, GASB’s proposed guidelines for service efforts and accomplishments reporting underscore its ongoing commitment to improving performance and data transparency, which are crucial for evaluating service efficiency and effectiveness (Sage & Sage, 2010.). These developments, in tandem with the advent of AI and tools like ChatGPT, present an unprecedented opportunity for local policymakers to harness enhanced data availability and analytical capabilities for more informed decision-making.

The rapid rise of AI has also raised concerns in the public sector, particularly regarding transparency. Recent work by Schiff, Schiff, and Pierson (2022) finds that “public value failures associated with artificial intelligence have significant negative impacts on citizens’ evaluation of public service.” These perceptions were primarily caused by citizens’ concerns about the lack of transparency in policy processes run by AI. This finding is consistent with Wenzelburger et al. (2024), who find that algorithms (critical to AI) are more accepted by citizens when they view the organization operating the algorithm more favorably (trust) and when the algorithm is of personal importance. Hence, financial managers must be given care to ensure transparency in AI processes. We thus have compelling reasons to understand better the utility of recent technological advances, specifically AI, for local government administrators. There is speculation about a range of possible outcomes from AI for budgeting and finance professionals. These possibilities range from AI replacing budgeting and finance professionals to having

limited utility, particularly for local government officials in smaller governments who lack the time and resources to invest in these technologies. This experiment offers a practical assessment of existing AI technologies to articulate the challenges and benefits associated with ChatGPT in its current form.

## **Research Design**

With a focus on small local governments, we selected Fort Calhoun, NE, as our unit of analysis. Fort Calhoun is in eastern Nebraska, north of Omaha. The city's population is just over 1,100, yet it is in the top quartile of cities in Nebraska at 102nd out of 579. Therefore, while considered by most a small city, Fort Calhoun surpasses many cities in size in this state and region. Offering a comprehensive range of municipal services such as fire and rescue, parks and recreation, economic development, water, and sewer, Fort Calhoun operates under a council-mayor form of government. Although lacking a finance director or manager, it has a clerk, treasurer, and an administrator referred to as a community coordinator. With a low unemployment rate of 2.9 percent, Fort Calhoun fits our profile as a small city with limited capacity to conduct the financial analyses we expect from most larger cities. Ultimately, by assessing the strengths and weaknesses of each AI tool, this study will provide a discussion on their relevance to financial tasks for local government practitioners.

The financial data used in this study were collected from the Nebraska Auditor of Public Accounts (2024). The data are consistently reported to the state and have been reviewed by the Auditor's office for compliance with property tax and debt limitation requirements. The data comprises property taxes, debt, expenditures, revenues, and property valuation. We selected the years 2001 to 2016 due to the accessibility of the financial data.

The next section presents a comprehensive evaluation of three distinct versions of ChatGPT, namely the ChatGPT-3.5 free version, ChatGPT-3.5 as a Python tutorial, and the advanced ChatGPT 4.0. Our assessment explores key aspects such as the capacity to handle varying data volumes, the accuracy of outputs generated, and the time taken to reach desired results. Time considerations encompass both the experimentation with prompts to achieve the expected output and the execution of codes, either within ChatGPT or on external platforms. Each version's performance is scrutinized across critical tasks, including descriptive analysis, fiscal condition analysis, forecasting, visualization, and policy recommendations.

## **Results**

### ChatGPT-3.5

ChatGPT-3.5's ability to internalize data is limited. When data was copied and pasted from an Excel spreadsheet into ChatGPT-3.5, it became evident that the program was constrained in processing information. Specifically, the program was limited to about nine columns and 16 rows of data.

Table 1. Descriptive Analysis, Valuation

<b>Descriptive Analysis</b>	<b>ChatGPT-3.5 Output</b>	<b>MS Excel Output</b>
Mean	\$54,852,713	\$56,417,414
Median	\$66,495,044	\$61,150,868
Minimum	\$40,507,996	\$40,507,996
Maximum	\$69,395,085	\$69,395,085
Standard Deviation	\$9,543,006	\$11,195,600

### *Descriptive Analysis*

Focusing on financial data from 2001-2016 for Fort Calhoun, NE, our descriptive analysis successfully provided an output for mean, median, minimum, maximum, and standard deviation, as shown in Table 1, with valuation as an example. However, subsequent error checking via Microsoft Excel revealed slight discrepancies in the populated mean, median, and standard deviation, whereas the minimum and maximum values were generally accurate.

Next, we requested ChatGPT-3.5 to identify patterns, fluctuations, or significant changes in the imported data. This prompt yielded the following valuation output: “The valuation of Washington County’s Fort Calhoun subdivision shows an overall increasing trend from 2000-2001 to 2015-2016, with some fluctuations along the way. There is a notable spike in valuation around 2005-2006 and a steady increase thereafter.”

### *Fiscal Condition Analysis*

For the fiscal condition analysis, we prompted ChatGPT with the instruction, “Please perform some fiscal condition analyses.” However, instead of executing any analysis, ChatGPT-3.5 outlined four possible options: “Relationship between Valuation and Net Cash Balance, Relationship between General Government Expenditure and Total Disbursements & Transfers, Debt Analysis, and Relationship between Total Property Tax Request and General Government Expenditure.” Only after further prompting with “Please conduct the four analyses as suggested” did the model proceed to conduct the specified fiscal analyses. Here is an example output of a fiscal condition analysis conducted by ChatGPT on the relationship between total property tax request and general government expenditure for Fort Calhoun, along with a summary of its findings:

- “Relationship between Total Property Tax Request and General Government Expenditure:
  - Calculating the correlation coefficient between total property tax request and general government expenditure:
- Correlation coefficient: 0.043
- The low correlation coefficient suggests a weak relationship between total property tax request and general government expenditure. It indicates that there is no significant association between the amount of property tax revenue generated and the level of government expenditure.”

Table 2. Financial Health Analysis, Net Cash Balance as a Percentage of General Government Expenditure

<b>Year</b>	<b>Percentage</b>
2000-2001	0.43%
2001-2002	0.81%
2002-2003	0.93%
2003-2004	1.50%
2004-2005	1.30%
2005-2006	1.46%
2006-2007	0.83%
2007-2008	2.17%
2008-2009	42.35%
2009-2010	7.13%
2010-2011	10.13%
2011-2012	33.67%
2012-2013	23.20%
2013-2014	26.64%
2014-2015	24.84%
2015-2016	19.66%

Overall, this proved to be useful as the output included correlation coefficients, directions, and strengths of the relationships. These results can help identify the degree to which two variables relate to each other, i.e., how one variable might co-vary with another. In this example of property tax and expenditures, a positive correlation could suggest that as property tax revenue increases, spending on government programs and services is likely to increase as well. However, because the strength of the correlation is weak, perhaps property tax has a limited influence on government spending decisions.

Similarly, ChatGPT-3.5 only generated descriptions of various financial health analyses – “debt-to-valuation ratio, net cash balance as a percentage of general government expenditure, property tax dependency, and debt service ratio” – without actually executing them. Again, only through additional prompting did ChatGPT-3.5 perform its suggested analyses with outputs including percentages and ranges over time.

An example output produced by ChatGPT on the net cash balance as a percentage of general government expenditures for Fort Calhoun with a summary of its findings is provided below and in Table 2.

- “Net Cash Balance as a Percentage of General Government Expenditure:
  - Calculate the net cash balance as a percentage of general government expenditure by dividing the net cash balance by general government expenditure and multiplying by 100:
- Using the provided data, the Net Cash Balance as a Percentage of General Government Expenditure for each year is as follows:
- The Net Cash Balance as a Percentage of General Government Expenditure represents the cash reserves relative to the government’s overall expenditure. The

values range from around 0.43% to 42.35% over the years, indicating varying levels of liquidity.”

### *Policy Recommendations*

Wanting to make use of the analysis, we prompted ChatGPT to provide some policy recommendations based on the two sets of financial condition analyses conducted: “Please provide some policy recommendation based on the two sets of financial condition analyses conducted.” The policy recommendations from the analysis included debt management, cash reserve management, revenue diversity, debt service planning, long-term financial planning, and regular monitoring and reporting. These policy recommendations seemed to be well thought-out, as seen through the concise, detailed recommendations that incorporated specific insights from the fiscal condition analyses done above. Below is an example of output produced by ChatGPT on Cash Reserve Management for Fort Calhoun:

- “Cash Reserve Management:
  - As indicated by the net cash balance as a percentage of general government expenditure, the subdivision has experienced varying levels of liquidity. It is recommended to maintain an adequate cash reserve that can cover a reasonable proportion of general government expenditure to ensure financial stability during times of unexpected expenses or revenue shortfalls.
- Suggestions for Improvement:
  - It is important to note that these recommendations should be considered in conjunction with a more detailed understanding of the specific circumstances, goals, and constraints of the Fort Calhoun subdivision. Additionally, involving relevant stakeholders and seeking expert financial advice can further enhance the effectiveness of these policy recommendations.”

### *Forecasting Data*

Our next task for ChatGPT-3.5 was to demonstrate its ability to provide simple forecasts for the imported variables of valuation, net cash balance, general government expenditures, total disbursements and Transfers, total property tax request, and total outstanding debt for the next 15 years using only data from the original dataset. To do so, the model used the average annual growth rate that it had calculated.

An example output of the forecasted Valuation values produced by ChatGPT for Fort Calhoun with some suggestions for improvement is provided below and in Table 3.

- “Here is the forecast for the next 15 years (from 2015-2016) using the average annual growth rates:
- Valuation Forecast:
  - Average Annual Growth Rate: 2.63%
- Please note that these forecasts are based solely on historical data trends and do not take into account any external factors or changes in economic conditions. They should be used as a rough estimation and further analysis with up-to-date

Table 3. Valuation Forecast

<b>Year</b>	<b>Valuation Forecast</b>
2016	\$71,283,454
2017	\$73,195,188
2018	\$75,180,356
2019	\$77,240,297
2020	\$79,376,445
2021	\$81,590,338
2022	\$83,883,515
2023	\$86,257,520
2024	\$88,713,903
2025	\$91,254,215
2026	\$93,880,009
2027	\$96,592,842
2028	\$99,394,380
2029	\$102,284,251
2030	\$105,300,094

- data and careful consideration of other factors is recommended for more accurate projections.”

### *Discussion of ChatGPT-3.5 Free Version*

ChatGPT-3.5 (free version) is limited in terms of data import and processing capacity. However, the output was generally useful and detailed when carefully prompted. Specifically, ChatGPT was effective in identifying fluctuations, general trends, and spikes in the data, which could be beneficial for analysts, as fluctuations or spikes in data can prompt a thorough review of data entry to ensure accuracy in the dataset. Furthermore, general trends in the data can provide analysts with a quick understanding of the overall direction of a city’s finances and identify potential data input errors. Although ChatGPT consistently provided suggestions for both sets of financial condition analyses, we found it necessary to explicitly prompt it to execute the analysis rather than solely describing the procedure. While ChatGPT-3.5 provided policy recommendations, it noted that such recommendations should be constructed with the involvement of relevant stakeholders and financial experts and with additional factors like circumstances, goals, and constraints of the city in mind. ChatGPT was also able to conduct a simple forecast using the average annual growth rate of imported data, though it noted the limitations associated with such a forecast.

### Using ChatGPT-3.5 to code in Python

We first asked ChatGPT to recommend a good AI program for graphing data, and Python was its first recommendation. Python is a free programming language that is widely used for a variety of purposes, including data analysis. Figuring out how to operate in Python, we asked ChatGPT how to get started with Python. The program recommended choosing an “Integrated Development Environment” (IDE), which is an application that runs Python code. The first IDE-



recommended program was called Jupyter Notebook. We asked ChatGPT to explain Jupyter Notebook and how to install the program. Getting to the stage of being able to type code into Jupyter Notebook took several hours. It included asking ChatGPT a host of questions, including how to use the program, a basic code to run in Jupyter Notebook, how to use the program for graphing data, and how to import data into Jupyter Notebook. From our perspective, while time-consuming and cumbersome, one of the benefits of Python is its ability to read several file types, including MS Excel, so there was no need to convert the existing dataset into a CSV file and risk problems associated with the conversion process. Having never run code before, let alone Python, presented some challenges and a good deal of time to get these data in operational form. Along the way, we ran into multiple coding errors, but fortunately, ChatGPT was able to help.

Once these data were loaded into Python, we asked ChatGPT how to run descriptives. We were given the code and answers to a host of coding errors, but ultimately, the program generated descriptives that included the mean, standard deviation, minimum, maximum, and quartiles for property valuation, total outstanding debt, net balances, property taxes, total revenues, and total expenditures. After another series of struggles with Python and ChatGPT, we were able to run simple graphs in Jupyter Notebook. While a bit crude, Python generated trend lines and box plots for our key variables. From there, we pivoted to generating financial ratios that included revenues to expenditures, cash balances relative to expenditures, and debt by valuation.

The real challenge began when we tried to learn how to forecast in Python. We typed in ChatGPT, “You are an analyst. How can I get Jupyter Notebook to generate 5-year forecasts?” To which ChatGPT responded: “To generate 5-year forecasts in Jupyter Notebook, you can use various forecasting techniques and libraries available in Python. Here’s a general guide on how to generate 5-year forecasts.” ChatGPT proceeded to describe a 6-step process that included how to import the necessary libraries for data analysis and forecasting (e.g., pandas, numpy, and forecasting tools such as statsmodels and scikit-learn), how to load the data, selecting the forecasting model, training the forecasting model, generating forecasts, and visualizing the forecasts.

For the experienced programmer, these steps probably make sense. Truth be told, we spent more than five hours trying to get Python to run a forecast. Even after watching YouTube videos, our efforts were unsuccessful. We got error message after error message, and even when we asked ChatGPT to help us fix the errors, we were unsuccessful in getting Python to generate the forecast. In summary, while highly effective and powerful, the idea that a local finance officer or local government officer in a smaller community will have the time and patience to learn coding from ChatGPT to conduct financial analyses in Python is unrealistic.

ChatGPT-4

In the ChatGPT-4 experiment, we aligned the input and prompts with those used in the 3.5 experiments. A key distinction between the two ChatGPT versions is that with ChatGPT-4, the model not only suggested Python code but could also execute it through a code interpreter plugin. Users could input data, and when prompted, the program would execute it directly within the chat. Code execution occurred in a secure environment, and both results and the code were displayed directly in the chat box. The user could also ask questions or request further analysis, and ChatGPT-4 would respond with the appropriate Python code and code execution. Overall, ChatGPT-4 was able to perform an interactive data analysis, which allowed for a deeper exploration of the data and provided more nuanced insights. Additionally, ChatGPT-4 was able to present data and calculation results in a more conversational style than ChatGPT-3.5.

Table 4. Descriptive Analysis, Valuation

Descriptive Analysis	ChatGPT-4
Count	16
Mean	56,417,410
Standard Deviation	11,195,600
Min	40,508,000
25%	44,084,260
50%	61,150,870
75%	66,527,510
Max	69,395,085

### *Descriptive Analysis*

Focusing on the same data, our descriptive analysis successfully provided output for mean, median, minimum, maximum, and standard deviation. Below is an example output of valuation from ChatGPT-4 of both a summary and table (see Table 4):

- “Valuation: The average valuation is approximately 56,417,410. The standard deviation is 11,195,600, which indicates a relatively large spread around the mean. The minimum valuation is 40,507,996 and the maximum valuation is 69,395,085”.

### *Visualization and Explanation*

A unique feature of ChatGPT-4 is that the visualizer plugin can provide visualizations along with a brief narrative of the results, as shown in the example of valuation over the years in Figure 1 below. This made the analysis more accessible and understandable, especially for non-expert users.

- “To identify patterns, fluctuations, or significant changes in the data, let’s examine the line plots for each variable over the years:
- Valuation:
  - The valuation of Washington County’s Fort Calhoun subdivision shows an overall increasing trend from 2000-2001 to 2015-2016, with some fluctuations along the way. There is a notable spike in valuation around 2005-2006 and a steady increase thereafter.”
- Whereas the code interpreting GPT 4 would use Python to plot the data and provide an interpretation as follows:
- “Valuation: There is a general trend of increasing valuation over the years. There is a significant increase from 2005-2006 to 2008-2009, after which the growth slows down.”

Figure 1. Valuation of Fort Calhoun, ChatGPT 4

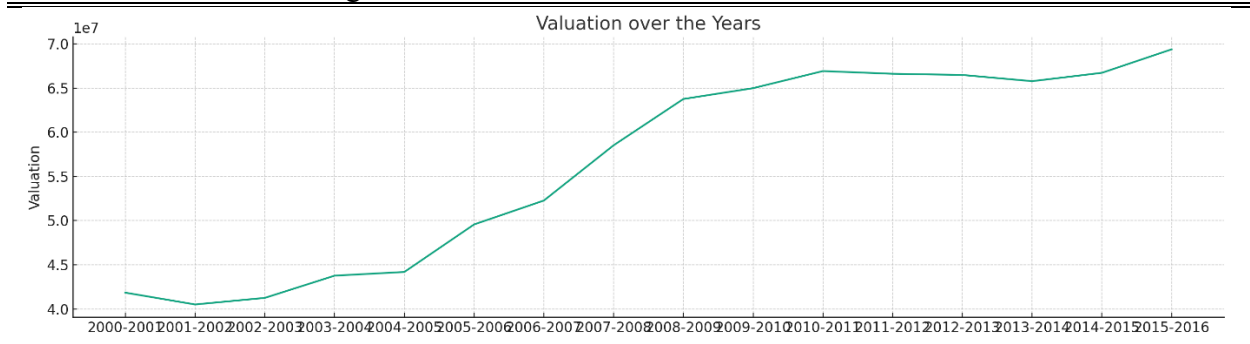
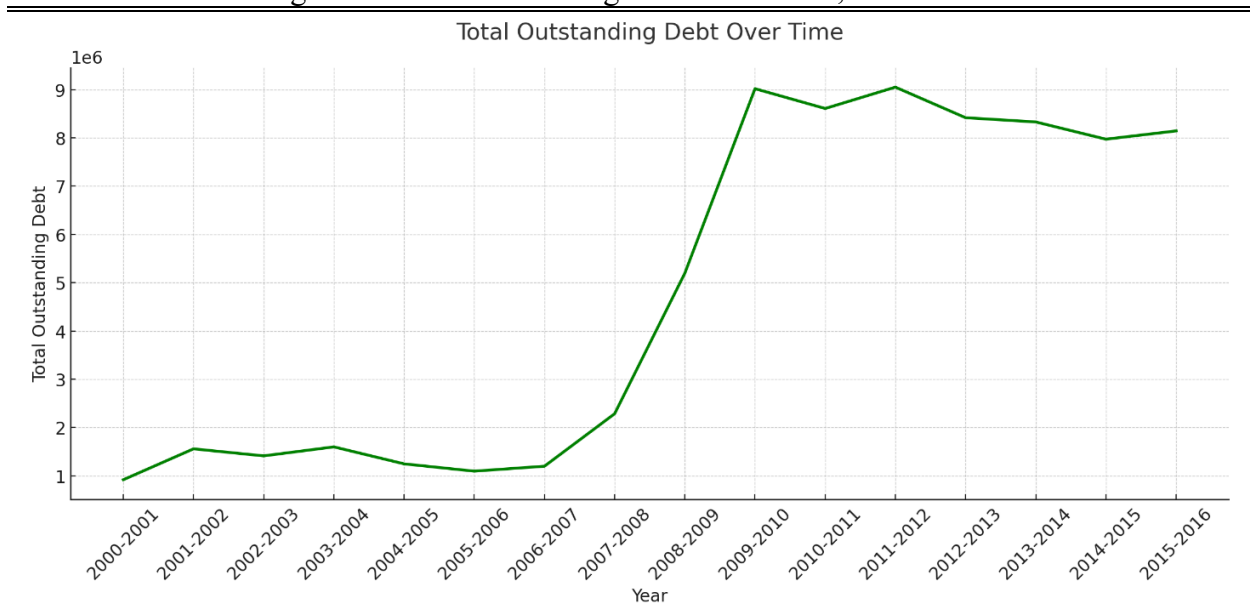


Figure 2. Total Outstanding Debt Over Time, ChatGPT 4



### Financial Condition Analysis

ChatGPT4 conducted a financial condition analysis using the following prompt: “Please conduct a financial condition analysis on the original dataset.” An example output with a visualization and description, along with some suggestions for improvement, is as follows (see also Figure 2):

- “There’s a significant spike in debt starting from around 2007-2008, reaching its peak in 2009-2010.
- After 2009-2010, the outstanding debt seems to have decreased slightly and then stabilized, with minor fluctuations in the subsequent years.
- The sharp increase in debt around 2007-2008 and its peak in 2009-2010 suggests that there might have been major investments, projects, or financial decisions made during that period. It would be essential to understand the reasons behind

this surge and how it aligns with the broader fiscal and strategic goals of Fort Calhoun.

- Overall, the financial health of a municipality depends on a combination of factors, including its ability to meet short-term obligations (liquidity), long-term financial stability (solvency), efficiency in government operations, and trends in valuation and debt. The analysis above provides a basic view of these aspects, but a more in-depth analysis could include additional factors such as revenue sources, expenditure commitments, demographic trends, and more.
- For a thorough assessment of financial health, it may also be helpful to compare these metrics with those of similar municipalities, or to benchmark against state or national averages.”

### *Policy Recommendations*

We prompted ChatGPT-4 to provide policy recommendations using the following prompt: “Please provide some policy recommendations based on the two sets of analyses conducted.” Here is an example output of a policy recommendation produced by ChatGPT-4 on Fort Calhoun:

- “Fort Calhoun has a high liquidity ratio, suggesting a strong short-term financial position. However, the high solvency ratio and increasing outstanding debt indicate potential long-term financial risks. Policies could be implemented to manage and reduce debt levels.
- Despite the high solvency ratio, the efficiency ratio is relatively low, suggesting effective management of disbursements and transfers. The municipality should continue to maintain this efficiency”.

These ChatGPT-4 recommendations were based on the analysis done, and the user could further interact with it to explore the provided recommendations, such as asking for more details or clarifications as needed. Notably, Chat GPT-4 moved through a series of ‘follow-ups,’ letting the user know of the limits of its recommendations and analyses based on the available data. An example follow-up provided by ChatGPT-4 is as such: “Remember, these recommendations are based on the data available and the analyses conducted. A more comprehensive policy recommendation would require a deeper understanding of the county’s specific context, needs, and constraints, as well as additional data and analyses.”

### *Forecasting*

Following this very brief and generic series of recommendations, we prompted ChatGPT-4 to do a forecast of the financial variables for the next 15 years. Here is an example output for the forecast done with the code interpreter plugin:

- “To provide a simple forecast for the next 15 years, we can use a method called linear regression. This method assumes that the future trend will continue in the same direction as the past trend. It should be noted that this is a simplified approach and doesn’t take into account potential changes in economic conditions, policy changes, or other factors that could impact these values in the future.

- Based on a linear regression model, below are the 15-year forecasted values for Fort Calhoun:
  - Forecasted Valuation: Increases from approximately \$66.6 million in 2016 to \$96.9 million in 2030.
  - Forecasted Net Cash Balance: Increases from approximately \$1.7 million in 2016 to \$2.5 million in 2030.
  - Forecasted General Government Expenditure: Increases from approximately \$577,077 in 2016 to \$815,138 in 2030.
  - Forecasted Total Disbursements & Transfers: Increases from approximately \$3.0 million in 2016 to \$4.3 million in 2030.
  - Forecasted Total Property Tax Request: Increases from approximately \$473,220 in 2016 to \$669,382 in 2030.
  - Forecasted Total Outstanding Debt: Increases from approximately \$5.6 million in 2016 to \$8.0 million in 2030.”

### *Discussion of ChatGPT-4*

Our exploration of financial data analysis tools engaged OpenAI’s ChatGPT-4 in a novel and interactive manner. ChatGPT -4’s integrated Jupyter notebook-like environment allowed us to perform sophisticated analysis on the Excel file containing Fort Calhoun’s financial data. For instance, ChatGPT-4 not only executed Python coding of data analysis but also offered detailed commentary at each stage of the output. Furthermore, the incorporation of a code interpreter within the chat interface made ChatGPT-4 highly beneficial to anyone, regardless of their coding background, even those with minimal coding proficiency. Also, the interactive nature of the interface permitted an iterative analysis process, fostering transparency through in-depth explanations at each step. However, the code interpreter’s limitations surfaced when complex and custom analyses demanded specific data types. Hence, even in the sandboxed local environment bounded by Python, a person with an innate understanding of the data being fed to the code interpreter must be present. This issue was mainly seen with the use of an unfiltered data spreadsheet. Hence, while the AI model could manage a wide range of analyses, specific cases or data types might fall outside its competency.

### **Challenges and Concerns**

One of the main challenges we encountered with the ChatGPT platform was the cost in terms of time, particularly when working with the free version of ChatGPT-3.5 alone and with Python. Specifically, in testing the capabilities of the free version, we noticed a steep learning curve where determining the optimal prompt proved to be time-consuming. On average, it took approximately five rewrites to achieve the desired results. As for ChatGPT-3.5 with Python, though this combination allowed for more flexibility and control, the main time cost was trying to decipher the optimal prompt for generating the intended code. Furthermore, the use of Python also required multiple troubleshooting and debugging efforts when errors occurred.

With ChatGPT-4, significant improvements in data processing enabled direct data file uploads, which significantly reduced time costs. However, despite the powerful capabilities of ChatGPT-4’s code interpreter, limitations persisted where a preference for “CSV” over “xlsx”

file types required additional data cleaning. Furthermore, the AI model lacked proficiency in handling specific complex analyses with simpler linear regression models for forecasting, which were chosen over more advanced options like Autoregressive Integrated Moving Average or machine learning models such as neural networks despite their being suggested. Still, accurate outputs were dependent on detailed or concise prompts, and ChatGPT-4 operated on a monthly subscription of \$20, unlike the free version of ChatGPT-3.5, with or without Python.

Retrospectively, here are a few lessons for the practitioner:

1. The quality of results depends on the quality of the data. While tools like Chat GPT and other AI models offer revolutionary analysis capabilities, they struggle with unreliable or unstructured data. Moreover, tools that aim to clean and correct data cannot help with such a situation at this point.
2. If practitioners lack knowledge and experience with the code interpreter's tools, they might overlook errors in calculations or thought processes that could potentially lead to inaccuracies. Practitioners need to be aware of their limitations when working with AI tools. Hence, we recommend integrating technical experts in the review process, especially given the public nature of the results. Specifically, local or national associations could offer specialized training programs that equip professionals with the necessary skills to effectively operate AI models accurately.

Overall, our experience with the ChatGPT platform highlights its powerful capabilities for generating outputs. The challenge, however, lies in the investment of time to understand and optimize prompts effectively. Transitioning to ChatGPT-4 improved data processing but introduced financial costs. We also observed better performance with a CSV data format and noted that regardless of which version was used, detailed prompts inevitably enhanced output quality. Hence, striking the right balance between time, cost, data preparation, and accurate and detailed prompts is essential for leveraging the ChatGPT platform's full potential in financial data analysis.

## Conclusion

This article introduces AI as a tool for analyzing Fort Calhoun's financial data through the following approaches: ChatGPT-3.5, ChatGPT-3.5 with Python, and ChatGPT-4. The tested AI capabilities cover descriptive analysis, fiscal health assessment, forecasting, visualization, and policy recommendations. ChatGPT's ability to perform these fiscal tasks is particularly valuable for smaller local governments with limited resources and capacity for financial analysis. Overall, ChatGPT-3.5 alone demonstrates promising potential for financial analysis in local government contexts, especially for smaller governments. Further advancements in the model's capabilities, when used with other programs or through the more powerful version of ChatGPT-4, AI could certainly benefit local governments in their financial decision-making processes.

A comparison between ChatGPT-4 and its predecessor, ChatGPT-3.5, reflects the strides made in AI. Unlike ChatGPT-3.5, which requires manual data input, ChatGPT-4 allows direct file uploads in the code interpreter. While ChatGPT-3.5 offers quick financial insights like averages, ranges, patterns, and even correlation analysis among variables, its output should be

treated as a general guide due to minor inaccuracies found in the output and its inability to process large datasets. Additionally, ChatGPT-3.5 lacks visualization capabilities. Hence, we recommend using the free version's outputs for a broad understanding rather than relying on them for practical applications or future research, especially in assessing the financial health of subdivisions like Fort Calhoun.

For Chatgpt-3.5 with Python, the output has the potential to be more sophisticated and accurate with the coding platform enabling visualizations for forecasting, descriptives, and fiscal analyses. However, this two-step approach requires someone with an adequate understanding of coding for effective utilization and troubleshooting. Hence, while ChatGPT-3.5 with Python holds the potential to achieve the fiscal analyses expected of a clerk or financial manager of a small city, the external execution of code poses a challenge for those lacking coding skills or an appropriate environment.

Contrastingly, ChatGPT-4 facilitates direct code execution within the chat interface, providing a seamless and accessible experience. However, challenges persist in terms of the data format and complexity of analysis as described. Nevertheless, the use of AI in city financial analyses holds promising implications for practice and policy, particularly in staff capacity enhancement. By automating data processing and analysis tasks, AI can enable staff to focus on higher-level decision-making and policy formulation. Furthermore, AI's capability to analyze historical data, forecast trends, and provide policy insights can lead to informed budget decisions and increased fiscal responsibility. Additionally, AI-generated policy recommendations that are tailored to community circumstances can offer policymakers valuable insights through pattern recognition in financial data for evidence-based decision-making.

We acknowledge that the findings of this study are limited to the ChatGPT platform, while other AI platforms like Bing AI and Anthropic Claude have been unexplored. Future research could thus compare different AI platforms to identify the most suitable one for specific financial analysis tasks. Also, while this study only explores the capabilities of ChatGPT through a single local government, future research could conduct a comparative analysis across multiple cities to deepen our understanding of AI. Additionally, as this study solely focuses on standard financial data, future research could explore different AI models for forecasting and analysis, incorporating economic data, market trends, and political or legislative changes. Finally, the study's emphasis on numerical data analysis could be broadened by investigating how AI can interpret financial text, such as Annual Comprehensive Financial Reports, Pension Valuation Reports, and Management Discussion and Analysis. This AI-driven text analysis could extract valuable insights, thereby enhancing accessibility and understanding for policymakers, stakeholders, and the community.

Engaging the community in AI usage is also a crucial aspect of future research. By leveraging AI for insights based on city financial data and reports, policymakers can foster transparency and involve citizens in the decision-making process. AI-powered financial tools can empower residents with accessible information, enabling participation in discussions on relevant fiscal matters. In conclusion, while this study highlights AI's benefits in local financial analysis, future research should explore diverse AI platforms, extend applications to financial text analysis, and how AI can actively engage the community for more informed decision-making and policy formulation in city finances.

## Disclosure Statement

The authors declare that there are no conflicts of interest that relate to the research, authorship, or publication of this article.

## References

- Andres, S. B. (2023). ChatGPT Resources for CPAs. *The CPA Journal*, 2023(May/June), 76-77.
- Baber, W. R., Beck, A., & Koester, A. (2024). Separation in the municipal debt market following GASB 34. *Accounting Review*. <https://doi.org/10.2139/ssrn.3663553>
- Ciancarini, P., Giancarlo, R., & Grimaudo, G. (2023). *Digital transformation in the public administrations: A guided tour for computer scientists*. <http://arxiv.org/abs/2305.05551>
- Coe, C. K., & Rivenbark, W. C. (2010). Implementing GASB 45: Recommended best practices in local government. *Public Budgeting & Finance*, 30(1), 71-81. <https://doi.org/10.1111/j.1540-5850.2010.00932.x>
- Dambra, M., Even-Tov, O., & Naughton, J. P. (2023). The economic consequences of GASB financial statement disclosure. *Journal of Accounting and Economics*, 75(2-3), 101555. <https://doi.org/10.1016/j.jacceco.2022.101555>
- Egodawale, M. H. A., Sedera, D., & Bui, V. (2022). A systematic review of digital transformation literature (2013 - 2021) and the development of an overarching a-priori model to guide future research. In *ACIS 2022 proceedings* (article 89). Association for Information Systems.
- Hinkley, S. (2023). *Technology in the public sector and the future of government work*. Center for Labor Research and Education, University of California Berkley.
- Kim, H. J., & Bretschneider, S. (2004). Local government information technology capacity: An exploratory theory. In R. H. Sprague (Ed.), *Proceedings of the 37th Hawaii international conference on system sciences* (pp. 1-10). IEEE. <https://doi.org/10.1109/HICSS.2004.1265310>
- Larson, S. E., & Overton, M. (2024). Modeling approach matters, but not as much as preprocessing: Comparison of machine learning and traditional revenue forecasting techniques. *Public Finance Journal*, 1(1), 20-39. <https://doi.org/10.59469/pfj.2024.8>
- Lowe, S. K. (2014). New GASB standards addressing public sector retirement systems: Impact on financial statements. *Proceedings of ASBBS*, 21(1), 435-442.
- Mission Square Research Institute. (2023). *State and local workforce: 2023 findings*. <https://research.missionsq.org/content/media/document/2023/05/2023workforce.pdf>
- National League of Cities. (2023). *Municipal workforce labor shortage*. <https://www.nlc.org/resource/improving-opportunities-and-boosting-economic-mobility/challenges/municipal-workforce-labor-shortage/>
- Nebraska Auditor of Public Accounts. (2024). *Basic budget data query*. <https://www.nebraska.gov/auditor/reports/index.cgi?budget=1>
- Pop! Automation. (2023, October 6). *How to use Chat GPT for financial analysis*. <https://www.popautomation.com/post/chatgpt-financial-analysis>
- Rivenbark, W. C., & Roenigk, D. J. (2011). Implementation of financial condition analysis in local government. *Public Administration Quarterly*, 35(2), 241-267.



- Sage, J. A., & Sage, L. G. (2010.). New guidelines for service efforts and accomplishments reporting in the United States. *International Journal of Public Information Systems*, 6(1), 111-119.
- Schiff, D. S., Schiff, K. J., & Pierson, P. (2022). Assessing public value failure in government adoption of artificial intelligence. *Public Administration*, 100(3), 653-673.  
<https://doi.org/10.1111/padm.12742>
- Srivastava, M., & Jain, K. (2023). Tracing evolution of management of technology discipline. *IEEE Transactions on Engineering Management*.  
<https://doi.org/10.1109/TEM.2023.3260861>
- Valle-Cruz, D., Gil-Garcia, J. R., & Fernandez-Cortez, V. (2020). Towards smarter public budgeting? Understanding the potential of artificial intelligence techniques to support decision making in government. In S. Eom & J. Lee (Eds.), *The 21st annual international conference on digital government research* (pp. 232-242). Association for Computing Machinery. <https://doi.org/10.1145/3396956.3396995>
- Walsh, M., & Nason, E. (2022). *Meeting demand for state and local public finance jobs*. Government Finance Officers Association.
- Wenzelburger G., Konig, P. D., Felfeli, J., & Achtziger, A. (2024). Algorithms in the public sector. Why context matters. *Public Administration*, 102(1), 40-60.  
<https://doi.org/10.1111/padm.12901>